Je-LKS

Journal of e-Learning and Knowledge Society
The Italian e-Learning Association Journal

sie-L
Società Italiana di e-Learning

NEW TRENDS, CHALLENGES AND PERSPECTIVES ON HEALTHCARE COGNITIVE COMPUTING: FROM INFORMATION EXTRACTION TO HEALTHCARE ANALYTICS

n. 1
2018
JANUARY

Je-LKS
Journal of e-Learning
and Knowledge Society

www.sie-l.it
www.je-lks.org

The Journal is issued online three times per year.

Je-LKS is an Open Access Online publication. This means that everybody can free access online to abstracts and full lenght articles.

Libraries or researchers, can subscribe a Reprint Service Subscription. The Reprint Service Subscription is on a yearly basis and you can receive the three printed year issues for a 121€ annual fee (including taxes and shipping expenses).

In order to join SIe-L:

- segreteria@sie-l.it
- Tel. +39 052 2522521

For more information visit www.sie-l.it

Registration at the Rome Court in the pipeline.

eISSN: 1971 - 8829 (online)

ISSN: 1826 - 6223 (paper)

Resp. dir. Aurelio Simone

To the authors:

paper can be addressed to:
www.je-lks.org

Je-LKS

Journal of e-Learning and Knowledge Society

The Italian e-Learning Association Journal

Vol. 14, n. 1, 2018

Je-LKS
Journal of e-Learning and Knowledge Society
The Italian e-Learning Association Journal

THE ISSUE

VOL. 14, N.1, 2018

by Nicola Villa

Je-LKS opens the new year with the issue dedicated to the theme **New Trends, Challenges and Perspectives on Healthcare Cognitive Computing: from Information Extraction to Healthcare Analytics**, edited by Mauro Coccoli (University of Genoa, Italy), Paolo Maresca (Federico II University of Naples, Italy) and Gabriella Tognola (Italian National Research Council, Italy).

All the paper of the special theme are introduced in the Editorial.

As usal Je-LKS publish also a selection of multidisciplinary paper accepted after a peer review procedure.

In particular **Francesca Pozzi**, **Donatella Persico** and **Luigi Sarti** in the paper *Evaluating Innovation Injection into Educational Contexts* propose an approach, called T&EAM (Technology & Event Acceptance Model), to evaluate the effectiveness of teacher training events together with indicators of the well-known Technology Acceptance Model, generally used to predict acceptance of a new technology.

Stefano Cacciamani, **Vittore Perrucci** and **Ahmad Khanlar** (*Conversational Functions for Knowledge Building: a study of an online course at University*) in their study want to explore two Specific Conversational Functions (SCF) and how they are used by the students in online courses.

Ronald Ochieng Ojino and **Luisa Mich** (*Mobile Applications in University Education: the case of Kenya*) examine the use of mobile applications in university education, focusing on Kenya.

Closes the issue the paper by **Marko Radovan** and **Mojca Perdih** (*Analysing Accessibility, Usability and Readability of Web-based Learning Materials – Case study of e-learning portals in Slovenia*) that attempts to compile a checklist that would identify the extent how web-based learning materials in Slovenia properly address the needs of people with dyslexia.

The next issue of Je-LKS (May 2018) will be dedicated to the theme ***Inital Teacher Education and Teacher Induction: what can we learn from International Practices?*** The call for paper, edited by Giuseppina Rita Mangione and Maria Chiara Pettenati (INDIRE), is currently open.

You can find all the information of the call and the free access to all the published paper on the journal's website **www.je-lks.org**.

Nicola Villa
Managing Editor
Journal of e-Learning and Knowledge Society

Je-LKS
Journal of e-Learning and Knowledge Society
The Italian e-Learning Association Journal

EDITORIAL

by Mauro Coccoli, Paolo Maresca, Gabriella Tognola

Focus on:

New trends, challenges and perspectives on healthcare cognitive computing: from information extraction to healthcare analytics

The focus of this special issue is cognitive computing in healthcare, due to the ever-increasing interest it is gaining for both research purposes and clinical applications. Indeed, cognitive computing is a challenging technology in many fields of application (Banavar, 2016) such as, e.g., medicine, education or economics (Coccoli *et al.*, 2016) especially for the management of huge quantities of information where cognitive computing techniques push applications based on the use of big data (Coccoli *et al.*, 2017). An unprecedented amount of data is made available from a heterogeneous variety of sources and this is true also in the case of health data, which can be exploited in many ways by means of sophisticated cognitive computing solutions and related technologies, such as, e.g., information extraction, natural language processing, and analytics. Also, from the point of view of programming they set challenging issues (see, e.g., Coccoli *et al.*, 2015). In fact, the amount of healthcare that is now available and, potentially useful to care teams, reached 150 Exabytes worldwide and about 80% of this huge volume of data is in an unstructured form, being thus somehow invisible to systems. Hence, it is clear that cognitive computing and data analytics are the two key factors we have for make use – at least partially – of such a big volume of data. This can lead to personalized health solutions and healthcare systems that are more reliable, effective and efficient also reducing their expenditures.

Healthcare will have a big impact on industry and research. However, this field, which seems to be a new era for our society, requires many scientific endeavours. Just to name a few, you need to create a hybrid and secure cloud to guarantee the security and confidentiality of health data, especially when smartphones or similar devices are used with specific app (see, e.g., Mazurczyk & Caviglione, 2015). Beside the cloud, you also need to consider novel architectures and data platforms that shall be different from the existing ones, because 90% of health and biomedical data are images and also because 80% of health data in the world is not available on the Web.

This special issue wants to review state-of-the-art of issues and solutions of cognitive computing, focusing also on the current challenges and perspectives and includes a heterogeneous collection of papers covering the following topics: information extraction in healthcare applications, semantic analysis in medicine, data analytics in healthcare, machine learning and cognitive computing, data architecture for healthcare, data platform for healthcare, hybrid cloud for healthcare.

The remainder of this number is organized as follows:

Baya Naouel Barigou, **Fatiha Barigou** and **Baghdad Atmani**, in their paper "*Handling Negation to Improve Information Retrieval from French Clinical Reports*" take into account the problem of information extraction from electronical medical reports. In particular, in the French context, they face the problem of identifying negations and improve performances in current automated information retrieval systems.

Luca Canensi, **Giorgio Leonardi**, **Stefania Montani** and **Paolo Terenziani**, in the paper "*A Context Aware Miner for Medical Processes*" propose a novel approach to medical process mining that operates in a context-aware fashion. In particular, they show on a set of critical examples how the proposed solution is able to cope with a wide range of issues.

Antonella Carbonaro, **Filippo Piccinini** and **Roberto Reda**, in their paper "*Integrating Heterogeneous Data of Healthcare Devices to enable Domain Data Management*" seek to face semantic interoperability among diverse IoT fitness and wellness devices. To this aim, they present a novel framework based on an ontology developed ad hoc for the specific healthcare domain of fitness and wellness applications and a mapping system based on the RML language to facilitate data integration and sharing among the devices.

Mauro Coccoli and **Paolo Maresca**, in their paper "*Adopting Cognitive Computing Solutions in Healthcare*" present the possible motivations to adopt cognitive computing based solutions in the field of healthcare and survey experiences already done, through an overview of challenging research topics that showcase the main results already achieved in this field, where both academics and industries are making big efforts to improve the performances of current systems and to propose novel solutions based on the profitable exploitation of big data.

Andrea Damiani *et al.*, in their paper "*Preliminary Data Analysis in Healthcare Multicentric Data Mining: a Privacy-preserving Distributed Approach*" highlight the importance of shared ontologies in the era of cognitive healthcare systems. They present a privacy-preserving distributed approach as a preliminary data analysis tool to identify possible compliance issues and heterogeneity from the agreed multi-institutional research protocol before training a clinical prediction model.

In conclusion, in this issue we will describe some healthcare scenarios in which the use of cognitive system is giving great benefits.

Mauro Coccoli - *University of Genoa (Italy)*
Paolo Maresca - *Federico II University of Naples (Italy)*
Gabriella Tognola - *Italian National Research Council (Italy)*

REFERENCES

Banavar, G.S. (2016), *Cognitive computing: from breakthroughs in the lab to applications on the field*, in Proceedings of the 2016 IEEE International Conference on Big Data (Big Data), Washington, DC, 1-1.

Coccoli, M., Maresca, P. and Stanganelli, L. (2016), *Cognitive computing in education*, in Journal of e-Learning and Knowledge Society, 12(2), 55-69.

Coccoli, M., Maresca, P. and Stanganelli, L. (2017), *The role of big data and cognitive computing in the learning process*, in Journal of Visual Languages & Computing, 38(1), 97-103.

Coccoli, M., Maresca, P., Stanganelli, L. and Guercio, A. (2015), *An experience of collaboration using a PaaS for the smarter university model*, in Journal of Visual Languages & Computing, 31(1), 275-282.

Mazurczyk, W. and Caviglione, L. (2015), *Steganography in modern smartphones and mitigation techniques*, IEEE Communications Surveys & Tutorials, 17 (1), 334-357.

Je-LKS
Journal of e-Learning and Knowledge Society
The Italian e-Learning Association Journal
PEER REVIEWED PAPERS
NEW TRENDS, CHALLENGES AND PERSPECTIVES ON HEALTHCARE COGNITIVE COMPUTING: FROM INFORMATION EXTRACTION TO HEALTHCARE ANALYTICS
Vol. 14, n.1, 2018
ISSN: 1826-6223 | eISSN: 1971-8829

HANDLING NEGATION TO IMPROVE INFORMATION RETRIEVAL FROM FRENCH CLINICAL REPORTS

Baya Naouel Barigou
Fatiha Barigou
Baghdad Atmani

Computer Science Department, University Of Oran 1 Ahmed Ben Bella, Algeria
barigounaouel@gmail.com, fatbarigou@gmail.com, atmani.baghdad@gmail.com

Keywords: French clinical reports, medical information retrieval, negation identification, natural language processing.

The aim of this work is to develop a framework to cope with the negative context in French clinical reports and assess the effect of negation identification on the performance of medical information retrieval. The proposed work significantly improves the performance of information retrieval done on French clinical reports where the precision improves by 10%.

for citations:
Barigou B., Barigou F., Atmani B. (2018), *Handling Negation to Improve Information Retrieval from French Clinical Reports,* Journal of e-Learning and Knowledge Society, v.14, n.1, 11-31. ISSN: 1826-6223, e-ISSN:1971-8829
DOI: 10.20368/1971-8829/1455

1 Introduction

Electronic medical reports (EMRs) generally written in free text are considered a valuable source of clinical data; These reports are often consulted to verify a diagnosis by looking for similar cases and also see the treatment that has been prescribed. Thus, information retrieval (IR) from EMR helps clinicians identifying those reports relevant to their information need and allows them making the best clinical decision at the point of care. However, development of an IR system has always been a challenge in the medical field. In fact, processing medical documents put forth greater difficulty due to the inherent language problems like synonymy, polysemy, negation and context. In the perspective of information retrieval, this work deals with the problem of negation in French clinical reports.

Traditional keywords-based IR has many shortcomings, especially in the medical domain where the quality of the results can be decreased. Suppose a physician wants to identify a set of patients with *bronchiolitis*. The system can run a keyword search like *bronchiolitis,* which would identify documents containing the term *bronchiolitis* as given in expression 1-a. This, however, will return also documents that report a patient without *bronchiolitis*, through expressions like those in (1 b-c):

1	a	« Le début remonte à l'âge de 3 mois marqué par l'apparition d'une 1ère épisode de bronchiolite traité par Célestène-salbutamol» (The beginning dates back to the age of 3 months marked by the appearance of the first episode of bronchiolitis treated with Celestene-salbutamol)
	b	« L'enfant ne présente aucun signe de bronchiolite» ("The child has no signs of bronchiolitis")
	c	« Bronchiolite négative» ("bronchiolitis history is negative")

This inappropriate use of retrieval based on keyword search results in detriments of system precision. To identify only patients with *bronchiolitis,* the system needs to filter out expressions that do not show that the patient has a *bronchiolitis.*

Negation is a complex expressive linguistic phenomenon studied in linguistics and philosophy (Morante & Sporleder, 2012). Negation changes the meaning of sentences by denying or rejecting a statement, transforming a positive sentence into a negative sentence.

Today, negation detection is not only an emerging task in natural language processing (NLP) but also a relevant task in various applications such as Information Retrieval (IR), opinion mining and sentiment analysis.

Negation has been widely studied in medical documents. However, most of the works that have addressed this problem concern the English language. There are works that have been concerned only with identification of negation (Chapman *et al.*, 2001, Mutalik *et al.*, 2001), other works have analyzed the effect of negation during the search for information (Limsopatham *et al.*, 2012, Koopman *et al.*, 2010). In the case of the French language, less work related to medical information retrieval has focused on negated terms found in EMR. Therefore, this research paper aims to develop a complete IR system that supports negation in French EMR to reduce noise and then improve the quality of research. The developed system handles negation during reports indexing step and also during research step.

2 Background

This section summarizes the most closely developments related to our work.

2.1 Negation in the French language

Long before invention of computer science, several studies were conducted to deal with negation from a linguistic point of view: syntactic, lexical, and even semantic. However, automatic processing of negation is a newly emerging research area. In linguistics, the negation is opposed to the affirmation. And whatever the language, the negation is always marked, that is to say, that an unmarked assertion is considered, by default, as affirmative. The linguistic status of negation is therefore not equivalent to that of affirmation. The grammatical nature of the morphemes used to express negation varies according to the languages and context of enunciation. For the French language, one can mention: adverbs (e.g. ne... pas, non), adjectives (e.g. aucun, nul), verbs (e.g. ignorer: do not know), pronouns (e.g., rien (nothing), personne (nobody)), prepositions (e.g., sans (without), sauf (except)), coordinating conjunctions (e.g., ni), conjunctions of subordination (e.g. sans (without)), prefixes (eg im, mé, ana) and substantives with an intrinsic negative semantic value (e.g., carence (deficiency)); Table 1 show some forms of negation for the French language.

Table 1
DIFFERENT FORMS OF NEGATION

Type	Description	Examples
Grammatical negation (GN)	GN is expressed by grammatical structures : sans, en l'absence de, pas de, manque de, aucun...(without, in the absence of, no, lack of, none...)	"le patient n'a pas de diabète" (The patient does not have diabetes) "aucun signe de pneumonie" (No signs of pneumonia) "absence d'anésthésie" (There is no anesthesia)
Lexical negation	Several words carry the notion of negation by using prefixes Some verbs can be equivalent to a negation	"Mécontent" (unhappy), "analphabète" (illeterate) "dans ce cas, on supprime le médicament" (in this case no drugs)
Semantic negation	Interpretation of the sentence has a negative meaning	"L'examen clinique est négatif" (Clinical examination is negative) "le sirop pour la toux est contre-indiqué" (cough syrup is contraindicated)

2.2 Negation in the medical field

In the medical field, negation is a common practice when writing medical reports. Doctors use negation to express absence of signs, symptoms, or negative results of a medical test. According to (Ballesteros *et al.*, 2012), about 95% of radiology reports contain a negation of the form "*no sign of*" (in French "aucun signe de") or "*rupture of*" (In French "rupture de") or an expression in this sense. Further, doctors are not limited to the syntax of negation as defined in natural language (French for our case), they use also others expressions not known in the syntax of the French language. In Table 2 we give an example of four medical expressions that have the same semantic interpretation (i.e. negative interpretation).

Table 2
EXAMPLE OF NEGATIVE EXPRESSION USED IN FRENCH MEDICAL REPORTS

Medical expression	Interpretation
CRP[1]: (-)	Le test CRP est négatif (CRP test is negative)
CRP: négative.	
CRP négatif.	
CRP -	

Scope of negation

The question of negation scope in a sentence is critical for information

[1] C-reactive protein

retrieval. A distinction is made between the sentence negation, or phrastic, or total negation, and the constituent or partial negation. Thus, depending on the structure of the sentence, the scope would be:

Nominal phrase (NP) subject + intransitive verb→ the verb is denied. (e.g." L'enfant ne **tousse** pas": the *child does not cough*)

NP + ne transitive verb + pas + de + object of the verb→ the object of the verb is denied (e.g. " l'enfant ne présente pas de **signes de pneumonie**": *the child does not show **signs of pneumonia***)

NP + ne+ transitive verb + pas + de + object of the verb+ [, de + object of the verb]* → all the object of the verb are denied (e.g. " l'enfant ne présente pas de **signes de pneumonie,** ***de diabète, de choloéstérol***": *The child has no signs of pneumonia, diabetes, cholesterol)*

2.3 Related work

The previous work has mainly focused on detecting negated terms from medical documents to remove them before indexing documents. In addition, most of the proposed tools are designed for English-language medical texts. For example, Chapman *et al.* (2001) proposed a rule-based algorithm called NegEx for detecting negated findings and diseases in English radiology reports. Based on regular expressions, the principle of NegEx is: negation triggers (also called negation signs) are annotated in the sentence and then scope of these triggers are defined. Evaluation of NegEx uses 1000 sentences taken from discharge summaries and indexed by physicians. They obtained a specificity of 94.5% and a sensitivity of 77.8% percent. Considering the impressive results of this tool and its simplicity, it has been used in several NLP applications (Mitchell *et al.*, 2004, Meystre & Haug, 2006) and adapted to several languages like Swedish (Skeppstedt, 2011), French (Deleger & Grouin. 2012), Spanish (Stricker, Iacobacci & Cotik, 2015) and Germany (Cotik *et al.*, 2016). Similarly, Mutalik *et al.* (2001) proposed Negfinder a tool for detecting negations in discharge summaries and surgical notes. After replacement of the concepts by their corresponding UMLS terms, negation signs are identified and then a context-free grammar, known as Look-Ahead Left-to-right Rightmost-derivation associates these markers with UMLS concepts. The specificity of Negfinder was 97.7 percent, and the sensitivity was 95.3 percent.

Agrawal *et al.* (2010) present a conditional random field model, designed to detect negation cues and their respective scopes. The authors used the publicly available BioScope corpus[2] to train the proposed model. Experiments achieved an overall F-measure of 98% for detecting cues and 95% for detecting scopes.

An empirical analysis of the effect of negation on the search for information

[2] http://www.inf.u-szeged.hu/rgai/bioscope

is carried out in (Koopman *et al.*, 2010). The result of this study shows that negation has no major impact on research. But, it must be pointed out that the proposed algorithm does not consider complex forms of negation, their tool is limited to identifying simple forms such as "no", "negative" and "no". The study by Limsopatham *et al.* (2012) gets different results; unlike the results of Koopman *et al.* (2010), the effectiveness of IR has been improved. The authors propose NegFlag which is based on NegEX to detect negations. To avoid incorrectly returning a document, in which a given positive query term appears in a negative context, they propose to register a term negated by concatenating the prefix "no" to it. NegFlag is evaluated on 34 topics from the TREC 2011 Medical Record track. Compared to a baseline system, which does not consider negative context, this approach yielded a 2.9% relative improvement in P@10.

The authors in (Díaz, 2013), modeled a negation detection system by combining two classifiers in series, the first classifier determines whether a word of a sentence is a sign of negation or not and its position in the sentence. For sentences that contain a sign of negation, the second classifier determines whether the rest of the words in these sentences are in the negative part or not. Naïve Bayes (NB), C4.5, Support Vector Machine (SVM) were used in the experiment with a collection of medical documents from the corpus Bioscope. The author did not specify the values of the results but indicated that the results were good. The work of (Costumero *et al.*, 2014) is an adaptation of the NegEx algorithm (Chapman *et al.*, 2001) for the Spanish language. The results obtained for the detection of negation were weak compared to those of English texts. The authors justify this weakness by the difference between the two languages (English and Spanish) and state that a simple translation of the rules is not enough. Hence, a thorough analysis of Spanish grammar is needed to improve the performance of the Spanish version of the NegEx algorithm. Another similar work carried out by Stricker, Iacobacci and Cotik (2015) adapted NegEx for the Spanish Language. In this study, NexEg shows to perform better.

Mehrabi *et al.* (2015) propose DEEPEN to detect negation using NegEx and by taking into account the dependency relationship between negation words and concepts within a sentence. Their aim was to decrease NegEx's false positives. DEEPEN was tested using EMR data from Indiana University (IU) and it was also evaluated on Mayo Clinic dataset. Recently, Elazhary (2017) proposed a rule-based tool called NegMiner to address some of the shortcomings of the NegEx algorithm (Chapman *et al.*, 2001). Unlike NegEx, NegMiner can deal with longer sentences containing contiguous and multiple negations by exploiting some basic syntactic and semantic information. Experimental results have shown the superiority of the mining results of the NegMiner in comparison to the simulated NegEx algorithm (Chapman *et al.*, 2001).

3 Proposed work

A complete information retrieval system is proposed below. As shown in Fig.1, it consists of two components; the index-building component and the document retrieval component.

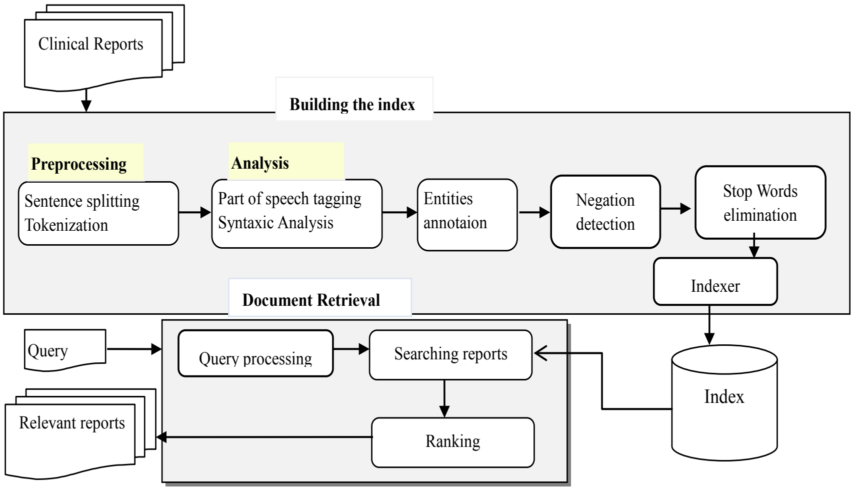

Fig.1 - Architecture of the proposed system

3.1 The index building

As indicated in Fig.1, the whole pipeline system works in five steps. We give the description of each step:

Preprocessing

The pre-processing step includes sentence boundary detection and tokenization. Decomposition of a report into a set of sentences is an important step for the detection of negation since our approach is based on the syntactic analysis of the sentences. The decomposition is mainly based on punctuation such as the dot "." the comma "," except that the majority of the sentences of the reports used in this study did not end with the point ". ", which required us to correct these reports before moving on to segmentation.

Linguistic analysis

This step includes part-of-speech (POS) tagging, lemmatization, and syntactic parsing. The Framework StanfordNLP version 3.6.0 is adopted where the French tagger library is used for POS tagging. Next, the results from POS tagging are fed to the Stanford parser to generate the syntactic tree.

Entities annotation

After linguistic analysis, terms annotation is carried out to distinguish between medical and non-medical terms. We need this annotation to weight the query terms during the research phase. The medical terms are more informative; as a result, they must have a higher weight than ordinary terms.

Medical entity identification is done using dictionary lookup. We compiled a list of commonly used entities (names of diseases, drugs, and symptom) from web sources like http://dictionnaire.doctissimo.fr and https://www.vidal.fr/ using the xidel[3] tool.

This resource is therefore used to annotate terms present in EMRs as medical terms if they are found in this resource, ordinary terms otherwise. Consider the following sentence as an example:

2A *Pas de notion de tuberculose pulmonaire mais présence de douleur thoracique (“No pulmonary tuberculosis but presence of chest pain “)*

pas :negation cue (“no” in english) considered as a stop word, it is not stored in the index
de : (“of”) stop word not stored in the index
notion: ordinary term
tuberculose: medical term
pulmonaire: medical term

In the above sentence of example 2-A, “*tuberculose*” (“*tuberculosis”*) is annotated as a medical term.

Negation detection and scope finding

Detection of negation and scope finding is typically divided into two subtasks. That is, first the trigger words of negation (negation cue) are detected and, thereafter, the task of determining the scope is carried out. Thus, a pattern-searching algorithm aimed to detect every word related to a negation cue analyzes the resulting syntactic tree. This step is performed based on a supplied list of models representing a negation format. A model of negation includes the syntactic expression of a sentence containing a sign of negation (e.g., “ne”, “ni”,

3 https://www.phpclasses.org/package/9136-PHP-Extract-information-from-HTML-using-the-Xidel-tool.html

"pas", "rien", "aucun", etc) and indicates the scope of this negation (the terms denied) in this sentence.

Detection of negation involves two steps. Candidate models are first identified; a vocabulary of trigger words of negation is compiled and words in the sentence are compared to this vocabulary in order to determine whether they are negation cues. If so, only models containing sign of negation similar to that of the sentence being analyzed are selected. In the second step, a match between the syntactic tree of the sentence and each candidate model is realized. This step makes it possible to identify the appropriate model that goes with the syntactic structure of the sentence.

Once the model corresponding to the sentence is identified, comes after the identification of negated terms (scope of negation sign). Once again, the two trees are compared in order to extract from the graph of the sentence the negated terms.

According to the French grammar, the scope of a negation trigger starts from the first word following the negation trigger and extends to the end of the sentence. Nevertheless, there are exceptions where this rule is not respected; for example, the use of words (such as "mais" (but) in example 2) causes a premature termination of the negation scope. In addition, the scope of some negation triggers is to the left of the trigger instead of the right Consider the following models as an example: [4][5][6]

3 A. (SENT (NP (ADV (**pas**[4])) (P (*)) (NC (*negation[5]*)) (AP (ADJ (*négation*)))))
B. (SENT (NP (ADV (**pas**))(P (*)) (NC (*negation*)) (PP (P (*)) (NP (*negation*))))mais NP)
C. (SENT (NP (*negation*)) (PUNC (*)) (NP (ADJ (**négatif**[6]))))

The regular expressions of example 3 are used to identify the negative diagnosis. The two first examples depict the syntactic patterns of sentences containing the negation sign "*pas*" ("not" in English). The last one contains the syntactic structure of a sentence containing the negation sign "*négatif*" ("negative" in English).

Using the model 3-B for the sentence of example 2-A, we can easily identify that the scope of "*pas*" which is a trigger word for negation, is "*notion de tuberculose pulmonaire*". So, "tuberculose" and "pulmonaire" are both denied. Note that the nominal phrase "*présence de douleur thoracique*" is not affected by negation because of the presence of the word "*mais*" ("*but*").

---

[4] trigger word for negation

[5] scope of negation

[6] trigger word for negation

Stop Words elimination

Unlike the usual indexing process, the step of eliminating empty words occurs after the negation detection step and not at the beginning of the process. Indeed, some signs of negation such as "ne (no), pas (not), sans (without), aucun (no) " are considered empty words, but these words play an important role in the negation detection phase. For this purpose, this filtering step is postponed after negation detection step.

Indexer

When building the index, we consider the following constraints:

1. Retrieve information from clinical reports that are written in the French language
2. Distinguish between medical terms and normal terms when ranking retrieved documents
3. Consider the terms that are denied in the weighting formulae of ranking

To deal with negation when retrieving documents as a response to a user query, the proposed index must satisfy the following conditions:

1. The index must record whether and at which position a term occurs in a document,
2. The index must record whether and at which position a negated term occurs in a document,
3. The index must record all information about a term "ti" necessary for its weighting:
    - Term Frequency (TF(ti, dj)): The total number of occurrences of the term "ti" in a document dj without considering whether it is denied or not.
    - Document Frequency (DF(ti)): The total number of documents containing term "ti".
    - Term Denied Frequency (TDF(ti, dj)): the number of times the term "ti" is negated in a document dj.
    - Document Denied Frequency (DDF(ti)): the number of documents containing the denied form of ti.

To deal with medical information retrieval, we have decided to distinguish between ordinary terms and medical terms during the retrieval. For that, in the case of noun terms, the index must record their type (medical term or ordinary term).

Thus, the core of the index consists of:

1. The dictionary represented by a hash table containing representative terms in documents. The dictionary records the root of the term, its DF, DDF, type (“M”: medical or “O”: Ordinary) and a pointer to its postings list.
2. The postings list: a linked list is associated with each term in the dictionary. Each posting is a docID with a list of positions, TF and TDF. When an occurrence of the term is negated, the position of this occurrence is also negated.

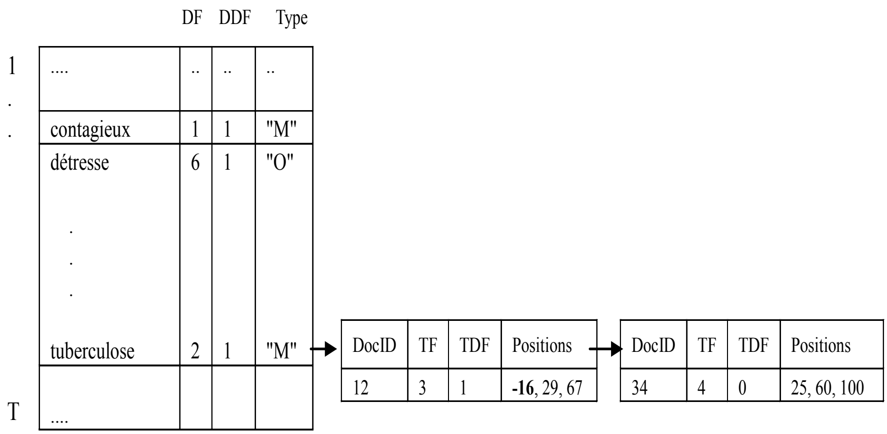

Fig.2 - Dictionary and Posting Lists structure

As an example, we have in Fig.2, the term “tuberculose” (“tuberculosis”) labeled as a medical term (Type = “M”), it is present in 2 reports (DF= 2) and negated in one report (DDF=1). In the report of ID 12, this term occurs three times (TF =3) where one occurrence is negated (TDF =1). The negated occurrence is at the position 16 (the position is prefixed with minus when the occurrence is negated).

In general, the structure used for the index is crucial for the efficiency of Information Retrieval (IR) systems. To speed up the retrieval of documents, we used Bitset vectors for representing Document-Term structure (DT), Document-Term Denied structure (DTD) and Term-Document structure (TD). The details about these three vectors are given in the following section. We consider “T” as the size of the dictionary (number of representative terms) and “N” as the number of documents in a collection.

Document-Term structure

The document-term structure is a list of Bit set vectors; DT(j); j=1 to N, where the j-th Bitset vector (noted DT(j)) corresponds to the jth document in the collection (its DocId from 1 to N) and the i-th Bit of each DT(j) represents the presence/absence of term “ti” in the j-th document. It is set to 1 if “ti” is present in this document, 0 otherwise.

Document-Term Denied structure

The document-term denied structure is like the Document-Term structure but here we focus on a document containing terms that are denied. The j-th Bitset vector (noted DTD(j)) corresponds to the jth document in the collection (its DocId from 1 to N) and the i-th Bit of each DTD(j) is set to 1 if all the occurrences of the i-th term are negated in this j-th document, otherwise 0.

Term-Document structure

The term-document structure is a list of Bit set vectors; TD(i); i=1 to M, where the i-th Bitset vector (noted TD(i)) corresponds to the i-th term of the dictionary (its row number in the array from 1 to M) and The j-th Bit of each TD(i) represents the presence/absence of the term “ti” in the document “dj” (j= 1 to N). It is set to 1 if the term is present in this document, 0 otherwise.

3.2 Retrieval step

Candidate documents searching

To speed up the process of retrieving documents and consider the presence of negated terms in these documents, we used both Bit set vectors’ DT and DNT to get only documents that match the query.

Also, the query “Q” is represented with a Bit set vector Q_v of size M where the i-th bit is set to 1 if the corresponding i-th index term is present in the query. It is set to 0 otherwise.

As shown in equation 1, identification of candidate documents consists in calculating for each document d_i(i = 1 to N) the Boolean vector *R(i)* using the Boolean vectors DT(i), DNT(i) and Q and the Boolean operations XOR and AND.

$$\mathrm{R(i)} = (\mathrm{DT(i)} \otimes \mathrm{DTN}) \wedge Q_v \qquad \text{eq.1}$$

According to equation 1, the document d(i) is a candidate document if $\exists\, \mathrm{j} \in [1, \mathrm{M}] / \mathrm{R(i, j)} = 1$, so documents containing the terms of the query

"Q" but which are denied are rejected by the system.

Scoring and ranking model

We used the state of the art term weighting models namely BM25 (Robertson *et al.*, 1996) to score and rank medical reports. For a given query q, the relevance score of document d, based on the BM25 term weighting model is expressed as

$$\mathrm{Score}_{BM25}(d, q) = \sum_{t \in q} IDF(t) * \frac{TF(t,d).(k+1)}{TF(t,d)+k.\left(1-b+b.\frac{L_d}{L_{ave}}\right)} \quad \text{eq.2}$$

where

- $IDF(t) = log\left[\frac{N}{DF(t)}\right] : DF(t)$ is the document frequency of term t and N is the number of documents in the collection;
- $TF(T_i,d)$ is the frequency of term T_i in document d
- k is the term frequency influence parameter which is set to 1.2 by default;
- b is the document normalization influence tuning parameter which has the default value of 0.75 (Robertson *et al.*, 1996);
- L_d and L_{ave} are respectively, the length of document d and the average document length for the whole collection.

In this work, we review the formula and added new information to reflect the presence of the negated terms in the documents and to associate weights to query terms. The relevance of a document for a given query is usually evaluated based on the weight of the query terms. This weight depends on the frequency of occurrence of the term in the document and, possibly, its frequency of occurrence in the collection. Compared to this classical model, taking account of the negation of terms in the weighting scheme consists in attributing weight to terms by considering their presence in a positive context or a negative context. Thus, the relevance of a document no longer depends only on the frequency with which the terms of the query appear, but also on the context (negative or positive) in which these terms appear in the document.

Thus, we propose a new BM25 weighting version for classifying relevant documents. This new version supports two important points:

- It considers that a medical term is more relevant than an ordinary term
- It considers the terms denied in the documents found by the system

The revised formula is given by the equation 3:

$$Score_{Mod\_BM25}(d,q) = \sum_{t \in q} w_t * IDF(t) * \frac{(TF(t,d) - NTF(t,d)).(k+1)}{TF(t,d) + k.\left(1 - b + b.\frac{L_d}{L_{ave}}\right)} \quad \text{eq.3}$$

- *NFT(t,d)* is the number of times the query term *t* is negated in document *d*
- w_t is the weight of term *t*
- $IDF(t) = \log \frac{N - (DF(t) - NDF(t)) + 0.5}{DF(t) + 0.5}$
- *NDF(t)* is the document frequency for which the term *t* is negated

We propose an extension of the probabilistic model (Maron *et al.*, 1960; Robertson *et al.*, 1976) that evaluates the relevance of a document for a given query through two probabilities: the probability of finding relevant information and finding irrelevant information. This extension exploits the fact that query terms can be negated in a medical report and in this case, this document must not be relevant to this query. Thus, a term will not have the same importance if it appears in a negative context or positive context. Document *d* is retrieved only if at least one appearance of the query term *t* in that document is not negated.

We consider that medical terms are more informative than the ordinary terms. For example, in "*détresse respiratoire*" ("respiratory distress") the medical term "respiratoire" is the most informative term to understand the noun phrase semantics while "détresse" ("*distress*") is an ordinary term that can be used in other noun phrases like "*détresse pulmonaire*" ("*pulmonary distress*"). For that, the weight w_t in the equation 3 takes into account the type of the query term. A medical term will have a greater weight than a non-medical term..

Let Q be a query and w_i be the weight of the ith term of the query Q.

We have then:

$$\sum_1^{|q|} w_i = |NMed| * w_{NMed} + |Med| * w_{Med} = Q_{NMed} + Q_{Med} = 1 \quad \text{eq4}$$

where

- |NMed| and |Med| are respectively the number of non-medical terms and the number of medical terms in the query Q (|NMed| + |Med| =|Q|)
- w_{NMed} and w_{Med} are respectively the weights of a non-medical term and a medical term.
- Q_{NMed} and W_{Med} are respectively the quantities of information provided by non-medical terms and medical terms in the query Q.

We can have three situations according to the values of |NMed| and |Med|:

1. All the terms of the query Q are non-medical terms, i.e. |Med|= 0 and |NMed| ≥1. In this situation we have :

$$Q_{NMed} = 1 \rightarrow \; w_{NMed} = \frac{1}{|NMed|} \qquad \text{eq.5}$$

2. All the terms of the query Q are medical terms, i.e. |NMed|= 0 and |Med| ≥1. In this situation, we have :

$$Q_{Med} = 1 \rightarrow \; w_{Med} = \frac{1}{|Med|} \qquad \text{eq.6}$$

3. The query **Q** contains **|Med|** medical terms and |**NMed|** non-medical terms**, i.e.** |NMed|> 0 and |Med|> 0

We calculate the weight of nom-medical terms following the equation 7

$$w_{NMed} = \frac{1}{|NMed|(1+\beta)} \qquad \text{eq.7}$$

We express the importance of the quantity of information Q_{Med} versus the quantity of information Q_{NMed} with the coefficient of importance β. To avoid assigning too much weight to non-medical terms we impose a constraint to ensure that a medical term will always have a higher weight than a non-medical term. We then have the following system of equations:

$$\begin{cases} \beta = \dfrac{Q_Med}{Q_{NM}ed} = \dfrac{|Med| * w_Med}{|NMed| * w_{NM}ed} \\ w_{NMed} < w_{Med} \\ \beta > 1 \end{cases} \qquad \text{eq.8}$$

The system admits a solution if $\mathbf{1} < \frac{|Med|}{|NMed|} < \boldsymbol{\beta}$ so it suffices to take a value $\boldsymbol{\beta} = \frac{|Med|}{|NMed|} + \mathbf{1}$

4 Experimental study and results

We present experiments to evaluate the performance of our proposed method. Our evaluation baseline is the BM25 model. Moreover, we compare our approach with the traditional keywords-based IR.

4.1 Document collection and Queries

Using a French corpus presents some challenges. To date, the pediatric hospital of Canastel[7] uses patients' records in the paper form and there is no real electronic benchmark. In this study, the pulmonary unit gives us access to 100 anonymous French reports for the period 2010 and 2011. We correct these reports (sentences which are not syntactically well-formed, and when there is lack of punctuation), enter, and save them in a text file format (txt). We also prepare a set of 21 queries that doctors may pose when faced with signs and symptoms of a medical condition of a new case. These queries are formulated with the help of a doctor of the same service. Priority is given to queries containing terms that are found in negated context.

4.2 Evaluation measures

We measure the retrieval performance of the proposed system in the term of Precision (P), Recall (R), and Mean Average Precision (MAP). The formulas of these measures are well defined in the book of (Manning *et al.*, 2008).

Precision (P) for a query Q is defined as the fraction of retrieved documents that are relevant (RDR) within the set of documents returned for Q (*RetD)*.

$$P = \frac{RDR}{RetD} \qquad \text{eq.9}$$

Recall (R) is the fraction of relevant documents that are retrieved by the system (RDR) over all relevant documents *RelD*.

$$R = \frac{RDR}{RelD} \qquad \text{eq.10}$$

Mean Average Precision (MAP) provides a single-figure measure of quality across recall levels. It is given by

$$MAP = \frac{1}{|NQ|} \sum_{j=1}^{|NQ|} AP_j \qquad \text{eq.11}$$

Where $|NQ|$ is the number of queries and AP_j is the average precision for a given query Q_j

$$AP_j = \frac{1}{nj} \sum_{i=1}^{nj} P@k \qquad \text{eq.12}$$

nj is the number of relevant documents for Q_j and *P@k* is the precision at rank position *k of the relevant document.* We used k=5.

[7] Located at west of Algeria

To evaluate the impact of the proposed index representation we have launched several experiments considering the following objectives:

- Impact of negation detection on Boolean retrieval
- Performance of the revised weighting measure BM25

4.3 Experiments results

Impact of negation detection on Boolean retrieval

For each query Qi (i= 1, 21), we calculate precision and recall. Then an average for each measure is calculated. The results obtained are shown in Table 3. As a first result, the Boolean search with the AND operator gives the best results in precision. It is clear that when the query is expressed in several terms, the precision of the system degrades with the OR operator. In addition, the most interesting of these results is that when the negation is active, the performance of the system still improves for queries that contain terms that are negated in some reports. These are, for example, requests Q1, Q4, Q7, Q9, Q15, and Q17. In these queries, considering the AND operator, the precision is much better when the negation is taken into account. On average, when the negation is not active, precision reaches the value of 67%. Moreover, when, the negation is active, precision increases with a rate of 8.58% to reach 71.77%. Fig.3 illustrates the performance of the system in the term of precision when using negation detection. As queries, we used only those that contain at least one term that is negated in clinical reports. We can easily see that the detection of negation greatly improves the results of research.

Table 3
IMPACT OF NEGATION DETECTION ON BOOLEAN RETRIEVAL; |Q| REPRESENTS THE NUMBER OF TERMS IN QUERY Q

Negation		non-active		non-active		active		active	
Operator		OR		AND		OR		AND	
	|Q|	P (%)	R (%)	P (%)	R (%)	P (%)	R (%)	P (%)	R (%)
Q1	3	18,75	100,00	65,00	100,00	19,05	100,00	80,00	100,00
Q3	2	3,45	100,00	18,18	100,00	3,45	100,00	18,18	100,00
Q4	1	66,67	100,00	66,67	100,00	100,00	100,00	100,00	100,00
Q5	1	77,78	100,00	77,78	100,00	77,78	100,00	77,78	100,00
Q6	2	45,45	100,00	45,45	100,00	45,45	100,00	45,45	100,00
Q7	4	37,50	100,00	62,16	100,00	44,44	100,00	98,46	100,00
Q8	3	38,89	100,00	66,67	100,00	38,89	100,00	66,67	100,00
Q9	2	6,41	100,00	71,43	100,00	6,94	100,00	83,33	100,00

Negation		non-active		non-active		active		active	
Operator		OR		AND		OR		AND	
	|Q|	P (%)	R (%)	P (%)	R (%)	P (%)	R (%)	P (%)	R (%)
Q10	3	15,00	100,00	37,50	100,00	16,67	100,00	37,50	100,00
Q11	1	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
Q12	2	11,29	100,00	53,85	100,00	11,29	100,00	53,85	100,00
Q13	2	34,78	100,00	68,57	100,00	34,78	100,00	68,57	100,00
Q14	2	16,13	96,15	66,67	96,15	16,13	96,15	66,67	96,15
Q15	2	6,67	100,00	66,67	100,00	45,00	100,00	100,00	100,00
Q16	2	12,50	100,00	12,50	100,00	25,00	100,00	25,00	100,00
Q17	2	6,15	100,00	80,00	100,00	7,41	100,00	100,00	100,00
Q18	2	5,26	100,00	33,33	100,00	5,26	100,00	33,33	100,00
Q19	2	5,88	100,00	100,00	95,83	5,88	100,00	100,00	95,83
Q20	5	3,57	100,00	100,00	100,00	3,57	100,00	100,00	100,00
Q21	1	71,43	100,00	71,43	100,00	80,65	100,00	80,65	100,00
Average		29,18	99,81	**63,19**	99,60	34,38	99,81	**71,77**	99,60

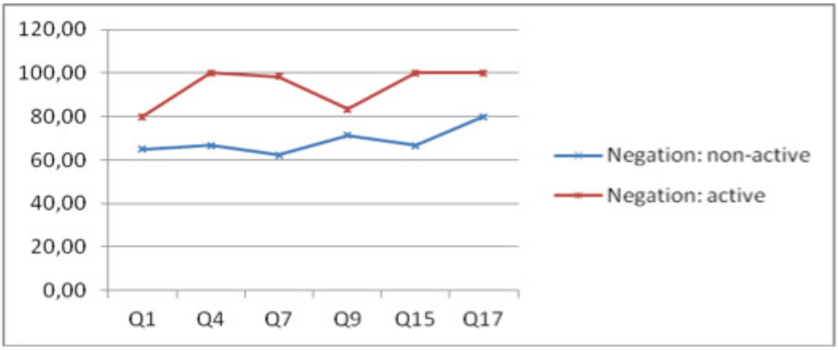

Fig. 3 - Performance of the system in term of precision when considering negation (here the AND operator is used for searching)

Measuring performance of the modified BM25

In the following experiments, we compute the average precision (AP) of each query then we compute the mean AP over the set of queries (MAP) by considering a search with the AND operator and taking into account the negation detection. The objective of these experiments is to analyze the performance of the modified BM25 measurement, which considers negated terms in reports. Results are presented in Fig.4.

Compared to BM25, it is clear that the new measure, which takes into account the terms negated in the reports, improves the quality of research. In addition, when the query includes medical terms the weighted Mod_BM25 gives the best results. As shown in Fig.4, the MAP of the modified BM25 (with/without term weighting) is the best.

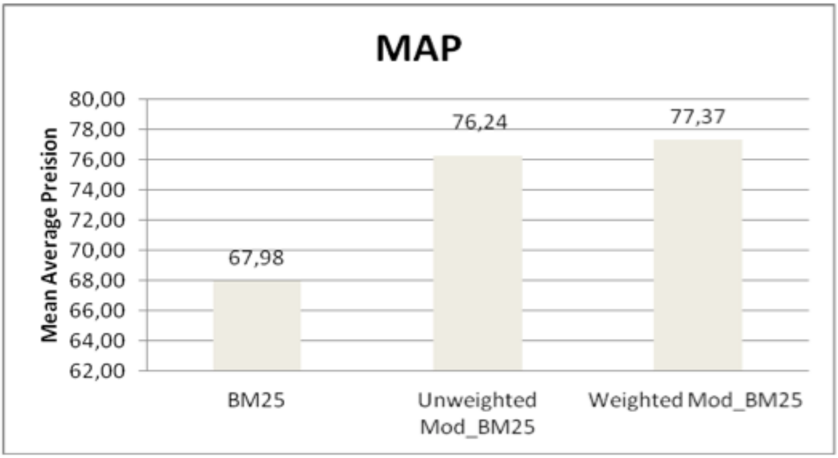

Fig.4 - The mean average precision with different term scoring

Conclusion

To improve the search quality of medical information retrieval, our solution is to introduce a negation detection mechanism to take into account this linguistic phenomenon not only during the indexing process but also during the search.

For this, we have designed and implemented an IR system based on a hybrid model (Boolean /probabilistic). The Boolean representation of the index allows to filter irrelevant documents i.e. documents containing no query terms or containing negated terms. The proposed adaptation of the BM25 formula takes into account the query terms that are negated in the reports and, finally the proposed term weighting considers the type of the query terms (medical or ordinary).

The results of our experiments, however modest they are, are still very encouraging. We achieved a significant improvement in precision compared to a system without detection of negation. These results confirm the importance of considering negation during the information retrieval process.

As prospects, there are still ways to explore and improvements to make. Therefore, several points can be considered:

- Enrich the base of negation models: A system that allows automatic generation of models from negative sentences annotated manually will ensure consistency between models and the detection of conflict situations or model redundancies because, with the evolution of the base models, the manual management of models will not be an easy task.

- Consider the structure of medical reports: medical reports have a well-defined structure that is characterized by a number of sections such as history section, clinical examination, and so on. A model that supports the search for answers at the different sections of the medical report will surely improve the accuracy of the results. This will allow the physician to search the information in well-defined sections of the document and thereby improve the accuracy of the results.

REFERENCES

Afzal, Z., Pons, E., Kang, N., Sturkenboom, M., Schuemie, M.J., Kors, J.A. (2014), *ContextD: an algorithm to identify contextual properties of medical terms in a Dutch clinical corpus BMC Bioinformatics*, 15, p. 373

Agarwal, S., Yu, H. (2010), *Biomedical negation scope detection with conditional random fields*. Journal of the American medical informatics association, 17(6):696–701.

Baldonado, M., Chang, C.K., Gravano, L., Paepcke, A. (1997), *The Stanford Digital Library Metadata Architecture*. *Int. J. Digit. Libr.* 1 108–121

Ballesteros, M., Francisco, V., Díaz, A., Herrera, J., & & Gervás. (2012), *Inferring the scope of negation in biomedical documents*. Computational Linguistics and Intelligent Text Processing, 363-375.

Bruce, K.B., Cardelli, L., Pierce, B. C. (1997), *Comparing Object Encodings*. Abadi, M., Ito, T. (eds.): Theoretical Aspects of Computer Software. Lecture Notes in Computer Science, Vol. 1281. Springer-Verlag, Berlin Heidelberg New York 415-438

Chapman, W., Bridewell, P., Hanbury, Cooper, G. F., Buchanan, B. G. (2001), *A simple algorithm for identifying negated findings and diseases in discharge summaries*. Journal of Biomedical Informatics, 34(5):301- 310.

Costumero, R., Lopez, F., Gonzalo-Martiın, C., Millan, M., & Menasalvas, E. (2014), *An Approach to Detect Negation on Medical Documents in Spanish*. Brain Informatics and Health, 366-375.

Díaz, N. P. (2013), *Detecting Negated and Uncertain Information in Biomedical and Review Texts*. The 9th International Conference on Recent Advances in Natural Language Processing, 45-50.

Elazhary, H. (2017), *NegMiner: An Automated Tool for Mining Negations from Electronic Narrative Medical Documents*. International Journal of Intelligent Systems and Applications, 9(4), pp.14-22.

Huang, Y., Lowe, H.J. (2007), *A novel hybrid approach to automated negation detection* *in clinical radiology reports*. Journal Am Med Inform Assoc, 14, pp. 304–311

Koopman B, Bruza P, Sitbon L, Lawley M (2010), *Analysis of the effect of negation on information retrieval of medical data*. Proceedings of the Fifteenth Australasian Document Computing Symposium (ADCS), Melbourne, Australia, pp 89–92

Limsopatham, N., Macdonald, C., McCreadie, R., Ounis, I. (2012), *Exploiting term dependence while handling negation in medical search*. Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval, 1065-1066. ACM.

Manning, D. C., Raghavan, P., and Schutze, H.(2008), *Introduction to Information Retrieval*. Cambridge Univ. Press, Cambridge

Maron, M. E, Kuhns, J. L. (1960), *On relevance, probabilistic indexing and information retrieval*. Journal of the ACM.

Mehrabi, S., Krishnan, A., Sohn, S., Roch, A.M., Schmidt, H., Kesterson, J., Beesley, C., Dexter, P., Schmidt, M., Liu, H., Palakal, M. (2015), *DEEPEN: A negation detection system for clinical text incorporating dependency relation into NegEx*. Journal of Biomedical Informatics, 54:213–219.

Michalewicz, Z. (1996), *Genetic Algorithms + Data Structures = Evolution Programs*. 3rd ed. Springer-Verlag, Berlin Heidelberg New York.

Mitchell, K. J., Becicha, M. J., Bermanc, J. J., Chapman, W., Gilbertsona, J., Guptaa, D., *et al.* (2004), *Implementation and Evaluation of a Negation Tagger in a Pipeline-based System for Information Extraction from Pathology Reports*. Medinfo, 663-7.

Meystre, S., Haug, P.J. (2006), *Natural language processing to extract medical problems from electronic clinical documents: performance evaluation*. J. Biomed. Inform. 39(6), 589-599.

Morante, R., Sporleder, C. (2012), *Modality and Negation: An Introduction to* the Special Issue. Computational Linguistics 38(2), 223-260.

Mutalik P., Deshpande A., Nadkarni P. (2001), *Use of general-purpose negation detection to augment concept indexing of medical documents*. Journal of the American Medical Informatics Association, 8(6), 598.

Robertson, S, Sparck Jones, K. (1976), *Relevance weighting of search terms*. Journal of the American Society for Information Science.

Roch, A.M., Mehrabi, S., Krishnan, A., Schmidt, H.E., Kesterson, J., Beesley, C., *et al.* (2014), *Automated pancreatic cyst screening using natural language processing: a new tool in the early detection of pancreatic cancer HPB*

van Leeuwen, J. (ed.) (1995), *Computer Science Today*. Recent Trends and Developments. Lecture Notes in Computer Science, Vol. 1000. Springer-Verlag, Berlin Heidelberg New York.

Je-LKS
Journal of e-Learning and Knowledge Society
The Italian e-Learning Association Journal

PEER REVIEWED PAPERS
NEW TRENDS, CHALLENGES AND PERSPECTIVES ON HEALTHCARE COGNITIVE COMPUTING: FROM INFORMATION EXTRACTION TO HEALTHCARE ANALYTICS
Vol. 14, n.1, 2018
ISSN: 1826-6223 | eISSN: 1971-8829

A CONTEXT-AWARE MINER FOR MEDICAL PROCESSES

Luca Canensi[1]
Giorgio Leonardi[2]
Stefania Montani[2]
Paolo Terenziani[2]

[1] Department of Computer Science, University of Torino, Italy
[2] DISIT, Computer Science Institute, University of Piemonte Orientale, Alessandria, Italy
canensi@di.unito.it, giorgio.leonardi,stefania.montani,paolo.terenziani@uniupo.it

Keywords: Process Mining, Medical Processes, Context

Medical process mining is gaining much attention in recent years, but the available mining algorithms can hardly cope with medical application peculiarities, that require to properly *contextualize* process patterns. Indeed, most approaches lose the connection between a mined pattern and the relevant portion of the input event log, and can have a limited precision, i.e., they can mine incorrect paths, never appearing in the input log traces. These issues can be very harmful in medical applications, where it is vital that mining results are reliable as much as possible, and properly reference the contextual information, in order to facilitate the work of physicians and hospital managers in guaranteeing the highest quality of service to patients. In this paper, we propose a novel approach to medical process mining that operates in a context-aware fashion. We show on a set of critical examples how our algorithm is able to cope with all the issues sketched above. In the future, we plan to test the approach on a real-world medical dataset, and to

for citations:
Canensi L., Leonardi G., Montani S., Terenziani P. (2018), *A Context Aware Miner for Medical Processes,* Journal of e-Learning and Knowledge Society, v.14, n.1, 33-44. ISSN: 1826-6223, e-ISSN:1971-8829
DOI: 10.20368/1971-8829/1453

extend the framework in order to support efficient and flexible trace querying as well.

1 Introduction

Process mining describes a family of a-posteriori analysis techniques that exploit the so-called *event log*, which records information about the sequences (*traces* henceforth) of events (i.e., activities) executed at a given organization. The most relevant and widely used process mining technique is discovery; process discovery takes as an input the event log and produces a process model, without using any a-priori information. The result is typically expressed in terms of a Petri Net, or some other process notation, often shown as a graph, in which nodes represent activities, and arcs provide control flow information. Medical process mining is a research field which is gaining attention in recent years (see, e.g., Mans *et al.*, 2009; 2008, Perimal-Lewis, 2012). This application domain indeed presents particular challenges and issues (Mans *et al.*, 2013), mostly related to the fact that different types of patients exhibit different characteristics, that the patient's state dynamically evolves (and is influenced by the medical activities executed on her/him), and that different hospital settings may have different resource constraints. All these peculiarities lead to the need to properly contextualize medical processes and/or process patterns. The currently available process mining solutions, ranging from commercial tools (offered, e.g., by Fujitsu Ltd and Celonis), to the open-source framework ProM (van Dongen *et al.*, 2005) (developed at the Eindhoven University of Technology), which represents the state of the art in process mining research, are not tailored to medical applications, and do not take into account contextualization needs. Specifically, despite some differences, many current algorithms, including *heuristic miner* (Weijters *et al.*, 2006) one of the most popular and widely used algorithms available in ProM, show important similarities (see Section 2.1), and have common limitations:

- they learn "context-free" patterns of processes;
- they can mine paths that do not correspond to any input trace in the log (i.e., they can have a limited *precision* (Buijs *et al.*, 2012);
- they do not explicitly relate the mined patterns to the log (in the sense that there is no explicit correspondence between mined patterns, and the traces in the log "supporting" them).

Such limitations are quite relevant in general, and very relevant in the medical domain. Concerning the first limitation, it is well known that, e.g., the same (set of) activities may produce different effects on patients, depending on the context (e.g., on the activities previously performed on the patients themselves). The impact of the second limitation is obvious and dramatic: if the miner precision is limited, in the sense that it may also learn a path that never appears in any input trace, this can be very harmful in medical applications, where it is vital that mining results are reliable as much as possible, in order to facilitate the work of physicians and hospital managers in guaranteeing the highest quality of service to patients. Indeed, the mined model is the input for quality assessment procedures, such as verification of conformance with respect to clinical guidelines, or performance measurement and bottleneck detection. All these procedures will provide an unreliable output, if played on an unreliable input. However, surprisingly, limited precision is a common limitation of many current miners (see Section 2.1.) The third limitation is less critical, but still significant. Indeed, maintaining an explicit link between mined patterns and the input traces matching such patterns, can be important not only to characterize contexts, but also to provide physicians with an evidence of the learned output, and also to provide support for retrieving traces corresponding to a given pattern.

In this paper, we propose an innovative approach that supports **context-aware** process mining, and overcomes all the above limitations.

2 A critical analysis of current approaches

As discussed in the Introduction, several different miners have been developed in the literature. However, many existing miners, including *alpha miner* (Van der Aalst & van Dongen, 2002) *fuzzy miner* (Gunther & Van der Aalst, 2007), *heuristic miner* (Weijters *et al.*, 2006) and the very recent inductive *tree miner* (Leemans *et al.*, 2013) show interesting commonalities. In this section, we will first illustrate common assumptions and methodological choices of the available mining approaches. Then, we will move to the discussion of the impact of these common issues on a set of examples. For the sake of clarity and brevity, we will refer to just one miner to illustrate the common characteristics of these literature approaches. Specifically, we will concentrate on heuristic miner. Indeed, heuristic miner can abstract from exceptional behavior and noise and, therefore, is suitable for many real-life logs, including medical ones. Moreover, it can mine the presence of cycles and of short distance and long distance dependencies, by means of dedicated procedures (see Section 2.1). Another advantage of this algorithm is that its default output graph can be easily converted to other types of process representation formalisms, including Petri Nets. These features make heuristic miner one of the most popular and widely used algorithms available in ProM, and a good reference for our comparisons.

2.1 Mining algorithms commonalities

Although different representation formalisms are adopted to model the mined processes (for instance, heuristic miner itself supports three different possible representation formalisms for the output), most mining algorithms are based on a common assumption:

Uniqueness assumption. Each event is unique in the output, so that each event appears at most once in the output of the process miner.

As a second commonality, the mining methodology is typically based on two steps (corresponding to steps 1 and 3 in Algorithm 1 below, which abstractly characterizes the behavior of heuristic miner on a set T of traces):

(i) a **de-structuring step**, in which immediate precedence between pairs of events are detected in the input traces, and

(ii) a **re-structuring step**, in which patterns of events are reconstructed, by combining the immediate precedence relation mined in the de-structuring step.

Algorithm 1: HM pseudocode

```
function HM(T) ;
(1) for each pair of events A and B do
  | search T to detect immediate precedence A → B and B → A
end
(2) Look for cycles of length one (same event repeated) or two (pair of events
repeated) ;
(3) for each triplet of events A,B, and C, such that A → B and A → C do
  | use formulae [6] to discriminate whether B and C are in AND (both of them
  | must be executed after A, in any order) or in XOR (exactly one of them
  | must be executed after A), and construct the output
end
(4) Look for "long-distance dependencies" between events, and add them to the
output.
```

The main critical issues in the algorithm are the following:

- C1 Given the uniqueness assumption, the de-structuring step (step 1 in Algorithm 1) evaluates the immediate precedence relations between events A and B looking at sequences “AB” and “BA” in the input traces, regardless of their position in the traces, and, thus, regardless of the context.
- C2 Given the fact that the de-structuring step ignores the context, and that the re-structuring step (step 3 in Algorithm 1) does not take into account the traces in the log, also the re-structuring step does not consider the context at all.
- C3 The uniqueness assumption imposes severe constraints on the re-structuring step.

In the following subsection, the critical impact of issues C1, C2 and C3 will be discussed on some easy examples. For the sake of concreteness, the examples will be processed using heuristic miner, and provided as Petri Nets, which are commonly assumed as an incontestable formalism to represent (the semantics of) processes.

2.2 Critical examples

To simplify the presentation, with no loss of generality, we suppose that each trace is prefixed with a distinguished starting symbol (*) and postfixed with another distinguished ending symbol (#).

Ex. 1 *Log content: 1000 equal traces: *ABCA#*

Although all the traces are equal, the approaches discussed in Section 2.1 cannot learn the (unique!) pattern. In this example, this is mainly due to the fact that the uniqueness assumption causes problems in the restructuring phase (critical issue C3): the two occurrences of the event A in the traces must correspond to a unique transition in the output Petri Net. Thus, a cycle (returning from C to A) must be introduced in the output. Notably, the output Petri Net also admits patterns like *A# or *ABCABCA#, which are not present in any of the input traces (see Figure 1(a)).

The pattern shown in this example represents one of the typical care processes for patients suffering from acute diseases. For instance, consider the diagnostic process of patients suffering from cerebral ischemia. After the symptom onset, the patient arrives at the emergency ward of the hospital. According to the latest medical guidelines for the treatment of stroke (Carlucci & Inzitari, 1996), it is recommended that the patient undergoes as soon as possible a computed tomography (CT), for: (1) the differential diagnosis between ischemic and hemorrhagic stroke and other cerebrovascular diseases, and (2) the identification of possible early signs of ischemic brain suffering. This action is indicated in Ex. 1 as action A = CT Execution. Subsequently, the patient should be evaluated by a trained neurologist, generating action B = Neurological Evaluation. Among the various additional investigations, which may be required depending on the type of patient, the guidelines suggest to always perform an electrocardiogram (ECG). In Ex. 1, this is referred to as action C = ECG Execution. Before transferring the patient to the Stroke Unit for further treatment, the guidelines recommend the repetition of the CT (within 48 hours), in order to obtain a better diagnostic and prognostic evaluation. The action A = CT Execution is therefore performed both as the first action of this process, and as the last one, thus generating the pattern of actions of this example.

In a similar manner, each of the following examples introduces situations which can be concretely instantiated as real clinical processes, or parts of them.

*Ex.2 Log content: 1000 equal traces: *ABCBA#*

This example highlights the limitations of "context-free"" approaches. Consider, in particular, critical issue C1 above: since the precedence relations are searched without considering the context (and on the basis of the uniqueness assumption), there is no way to decide whether A precedes B (100% of traces, considering the first part of the traces) or B precedes A (100% of traces, but considering the last part of the traces), and analogously for the relation between B and C. As a consequence, two separate Petri Nets are learned: the first contains only the pattern *A#, while the second contains a loop composed by C and B (see Figure 1(b)). By using heuristic miner without the "all event connected option", a different model is learned (see Figure 1(c)), where it is possible to replay the input trace, but many other never observed behaviors can be generated as well.

*Ex.3 Log content: 500 traces *AD# and 500 traces *DB#*

Once again, the combination of a "context-free" analysis with the uniqueness assumption causes undesired effects. In this example, two different patterns should be mined. However, during the de-structuring phase, no distinction is made between D in the first type of traces (i.e., D in the context of being preceded by *A) and D in the second type of traces (i.e., D in the context of being preceded by *). Thus, the re-structuring phase (see critical issue C2) generates a "merging" between the two different patterns (due to the uniqueness of D), providing the Petri Net in Figure 1(d) as output. Notably, besides the correct patterns *AD# and *DB#, the output Petri Net also models the patterns *D# and ADB#, which do not correspond to any trace in the input log.

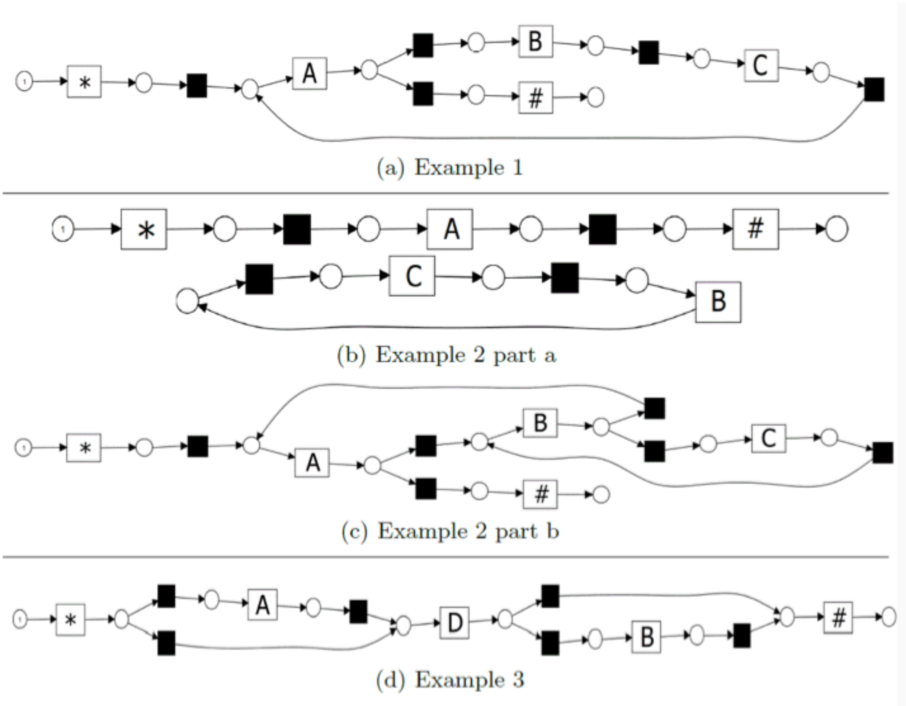

Fig. 1 - Models mined by heuristic miner referring to the examples of Section 2.2

3 Context-aware process mining

In order to overcome the limitations illustrated in the previous section, we propose an innovative approach to (medical) process mining, which is based on quite a different philosophy: in our methodology, process mining heavily takes into account contextual information, both in the data structure (the output graph), and in the mining algorithm. Our approach is based on three main ideas:

- in the data structure, we **remove the uniqueness assumption**: the same event may appear more than once in the output graph, to model the fact that it may occur in different contexts;
- in the data structure, and in the mining algorithm, we explicitly maintain the context, i.e., the set of log traces that *support* a given path in the mined graph;
- in the mining algorithm, we take advantage of the ordering of events in the log traces (which represents, indeed, the context in which each single event occurs) in order to directly mine the output graph, **without distinguishing between a de-structuring and a re-structuring phase**.

In the following, we present our approach and its application to the previously discussed critical examples.

3.1 Data structure and algorithm

Our mining algorithm takes in input a log (set of traces). The log is represented by a matrix, in which each row is a trace. It outputs the mined process as an acyclic directed graph, called a "process graph", in which nodes represent events (i.e., activities), and arcs represent a precedence relation between them. More precisely, in our model, each node is represented as a pair $< P; T >$:

- *P* denotes a (possibly unary) set of events; events in the same node are in *AND* relation, or, more properly, may occur in any order (with respect to each other). Note that, in such a way, each path from the starting node of the graph to a given node N denotes a set of possible process patterns (called *support patterns* of N henceforth), obtained by following the order represented by the arcs in the path to visit the process graph, and ordering in each possible way the events in each node (for instance, the path *{A,B}->{C}* represents the support patterns "ABC" and "BAC").
- *T* represents the context, i.e., a set of references to all and only those traces in the log which exactly match one of the patterns in *P* (called *support traces* henceforth).

Our mining algorithm operates in two steps. The first step (see Algorithm 2 below) is the core of our approach, and builds a tree of nodes described as above.

Algorithm 2: Mining pseudocode

```
function Build-Tree(index,< P,T >) ;
nextP ← getNext(index+1, T) ;
if nextP not empty then
    nextEvents ← XORvsAND (nextP, T) ;
    for each node < P′,T′ > ∈ nextEvents do
        AppendSon(< P′,T′ >,< P,T >) ;
        Build-Tree(index + |P′|,< P′,T′ >) ;
    end
end
```

Build-Tree in Algorithm 2 takes in input a variable *index*, representing a given position in the traces (i.e., a column in the input matrix), and a node. Initially, it is called on the first position, and on the root of the tree (which is a "dummy" node, corresponding to the * event; thus, initially, index=0, P=*

and T is the set of all the traces). The function *getNext* simply inspects the traces in *T* to find all possible next events (in the context *T*). On the basis of the current context (*support traces*) *T*, the function *XORvsAND* applies the formulae described in appendix A to identify which events are in AND and which are in XOR relation. The output of such a function is a set of nodes < *P', T'* >, one for each maximal set of events to be AND-ed. Note that, for each one of such sets *P'*, the corresponding set *T'* of support traces is also computed, on the basis of the current context *T*. Finally, each new node is appended in the output tree (function *AppendSon*), and *Build-Tree* is recursively applied to each node (with the parameter *index* properly set). The second step is a simple transformation of the tree into an acyclic directed graph, obtained by merging, starting from the leaves nodes, those paths in the tree whose *support pattern* postfixes are identical. Notably, our mining algorithm explicitly manages the context, focusing at each step on the proper *support traces*, and always taking into account the global ordering of events in the traces to build the graph. As a consequence, differently from the heuristic miner algorithm shown in Algorithm 1, we do not need any additional step to cope with long-distance dependencies (we have them "for free"). Analogously, we "naturally" cope with cycles by simply unfolding them. Thus, our algorithm already directly copes with *cycles* of any *length* (and no additional ad-hoc procedure for cycles is needed, differently from the heuristic miner algorithm).

3.2 Examples

In Figure 2, we show the output of our miner, applied to the examples Ex.1-Ex.3 above. Support traces are not reported in the figure, but are part of the output itself. As it can be observed, the critical situations discussed in Section 2.2 are all correctly managed by our approach. It is also worth noting that our process graphs could be easily converted into Petri Nets.

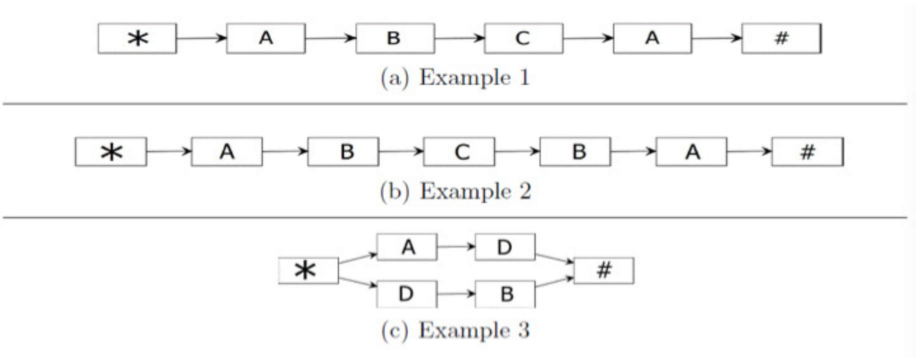

Fig. 2 - Models mined by our miner referring to the examples of Section 2.2

Conclusions

In this paper, we have introduced a novel process mining algorithm, able to overcome the limitations of many current approaches available in the literature. Specifically, our algorithm:

- learns "context-aware" patterns of processes;
- has a high precision, since it provides patterns that always correspond to input traces in the log;
- explicitly relates the mined patterns to the traces in the log "supporting" them.

Our approach properly deals with critical situations that may occur in practical application domains, as illustrated by the examples in section 2.2. These characteristics make the algorithm particularly well-suited for medical applications, where it is vital that mining results are reliable as much as possible, in order to facilitate the work of physicians and hospital managers in guaranteeing the highest quality of service to patients. In the future, we plan to test the approach on a real-world medical dataset, taken from the stroke patient management domain.

From the methodological viewpoint, we also aim at extending the framework, in order to support efficient and flexible trace querying. Indeed, the model we mine maintains an explicit link between mined patterns and the input traces supporting them, and can thus be seen as an indexing structure, well suited to quickly retrieve traces corresponding to the pattern at hand.

REFERENCES

Mans R., Schonenberg H., Song M., Van der Aalst W., Bakker P. (2009), *Application of process mining in healthcare - a case study in a Dutch hospital*, in: Fred A., Filipe J., Gamboa H. (eds), Proc. BEST, Communications in Computer and Information Science (25), 425-438, Springer.

Mans R., Schonenberg H., Leonardi G., Panzarasa S., Cavallini A., Quaglini S., Van der Aalst W (2008), *Process mining techniques: an application to stroke care*, in: Andersen S., Klein G., Schulz S., Aarts J. (eds), Proc. MIE, Studies in Health Technology and Informatics (136), 573-578. IOS Press.

Perimal-Lewis L., Qin S., Thompson C., Hakendorf P. (2012), *Gaining insight from patient journey data using a process-oriented analysis approach*, in: Butler-Henderson K., Gray K. (eds), Proc. Workshop HIKM, CRPIT (129), 5966, Australian Computer Society.

Mans R., Van der Aalst W., Vanwersch R., Moleman A. (2013), *Process mining in healthcare: data challenges when answering frequently posed questions*, in: Lenz

R., Miksch S., Peleg M., Reichert M., Riano D., ten Teije A. (eds),, ProHealth/KR4HC, LNCS (7738) 140-153. Springer.

van Dongen B., Alves De Medeiros A., Verbeek V., Weijters A., Van der Aalst W. (2005), *The proM framework: a new era in process mining tool support*, in: Ciardo G., Darondeau P. (eds), Knowledge Management and its Integrative Elements, 444-454, Springer.

Weijters .A., Van der Aalst W., Alves de Medeiros A. (2006), *Process Mining with the Heuristic Miner Algorithm*, WP 166, Eindhoven University of Technology.

Buijs J., van Dongen B., Van der Aalst W. (2012), *On the role of fitness, precision, generalization and simplicity in process discovery*, in: On the Move to Meaningful Internet Systems, OTM, 305-322. Springer.

Van der Aalst W., van Dongen B. (2002), *Discovering workflow performance models from timed logs*, in: Han Y., Tai S., Wikarski D. (eds), Proc. EDCIS 2002, LNCS (2480), 45-63, Springer.

Gunther G, Van der Aalst W. (2007), *Fuzzy mining - adaptive process simplification based on multi-perspective metrics*, in: Alonso G., Dadam P., Rosemann M. (eds), Proc. BPM LNCS (4714) 328-343, Springer.

Leemans S., Fahland D., Van der Aalst W. (2013), *Discovering block-structured process models from event logs containing infrequent behaviour*, in: Lohmann N., Song M., Wohed P. (eds), Proc. BPM Workshops, LNBIP (171) 66-78, Springer.

Carmona J., Cortadella J., Kishinevsky M. (2008), *A Region-Based Algorithm for Discovering Petri Nets from Event Logs*, Proc. BPM, 358-373, Springer.

Carlucci G., Inzitari D., (1996), *Italian stroke guidelines (spread): evidence and clinical practice*. Neurological Sciences (27) S225-S227.

APPENDIX: XOR VS AND FORMULAE

In Algorithm 2, after finding the set of successors *nextP* of the considered set of events P, we focus on the discovery of the relation between them. In order to do this, we calculate the *dependency frequency* between every event pairs < A;B >

in *nextP* _ *nextP* :

$$A \rightarrow B = \frac{1}{2}\left(\frac{|A > B|}{\sum_{X \in traces} |A > X|} + \frac{|A > B|}{\sum_{Y \in traces} |Y > B|}\right) \quad (1)$$

where $|A > B|$ is the number of traces in which A is immediately followed by B, and $|A > X|$ is the number of traces in which A is immediately followed by some event X, $|Y > B|$ is the number of traces in which B is immediately preceded by some event Y. After evaluating the dependency frequency value *A -> B* and *B -> A*, we can have the following possible situations:

- if both the values are below a given threshold, this means that A and B rarely appear in the same trace, therefore they are in XOR relation;

- if *A -> B* is above the threshold and *B -> A* is below, then A precedes B, and, viceversa;
- if both the values are above the threshold, then A and B are in AND (any-order) relation.

Je-LKS
Journal of e-Learning and Knowledge Society
The Italian e-Learning Association Journal
PEER REVIEWED PAPERS
NEW TRENDS, CHALLENGES AND PERSPECTIVES ON HEALTHCARE COGNITIVE COMPUTING: FROM INFORMATION EXTRACTION TO HEALTHCARE ANALYTICS
Vol. 14, n.1, 2018
ISSN: 1826-6223 | eISSN: 1971-8829

INTEGRATING HETEROGENEOUS DATA OF HEALTHCARE DEVICES TO ENABLE DOMAIN DATA MANAGEMENT

Antonella Carbonaro[1]
Filippo Piccinini[2]
Roberto Reda[3]

[1] Department of Computer Science and Engineering, University of Bologna, Italy - antonella.carbonaro@unibo.it
[2] Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori (IRST) S.r.l., IRCCS, Oncology Research Hospital, Meldola (FC), Italy - filippo.piccinini@irst.emr.it
[3] Master's Degree in Computer Science and Engineering, University of Bologna, Italy - roberto.reda@unibo.it

Keywords: Internet of Things; Cognitive computing; Healthcare Data; Ontologies; Semantic Web Technologies

The growth of data produced for example by IoT devices has playing a major role in developing healthcare applications able to handle vast amount of information. The challenge lies in representing volumes of data, integrating and understanding their various formats and sources. Cognitive computing systems offer promise for analysing, accessing, integrating, and investigating data in order to improve outcomes across many domains, including healthcare. This paper presents an ontology-based system for the eHealth domain. It provides semantic interoperability among heterogeneous IoT fitness and wellness devices and facilitates data integration and sharing. The novelty of the proposed approach lies in exploiting semantic web technologies to explicitly describe the meaning of sensor data and define a

for citations:
Carbonaro A., Piccinini F., Reda R. (2018), *Integrating Heterogeneous Data of Healthcare Devices to enable Domain Data Management,* Journal of e-Learning and Knowledge Society, v.14, n.1, 45-56. ISSN: 1826-6223, e-ISSN:1971-8829
DOI: 10.20368/1971-8829/1450

common communication strategy for information representation and exchange.

1 Introduction

The term eHealth refers to systems which leverage information and communications technologies to support healthcare and health-related fields including wellness and fitness industry.

Due to the ever-increasing usage of mobile health apps and wearable devices in our day-to-day lives, Internet of Things (IoT) technology is becoming one of the most popular trends for assisted eHealth (Islam *et al.*, 2015; Kim *et al.*, 2014). For instance, wrist-worn fitness bands such as Fitbit or Jawbone UP are wellness devices which can monitor 24/7 users' physical activity and health by keeping track of steps taken, distance travelled, stairs climbed and sleep hours. More advanced wearable fitness devices such as smartwatches, along with a GPS for outdoor sport tracking include also sensors for keeping track of basic physiological parameters of the wearer with an emphasis on heart rate and body temperature. Consumer IoT fitness trackers can even be used to monitor patient with health problems, for instance cardiologic and oncologic patients (Mendoza *et al.*, 2017).

Nowadays, thanks to the aforementioned smart wearable devices we can revolutionize the way we manage our physical well-being and training sessions by analyzing the large volumes of structured and unstructured produced health data.

However, collected data are often stored and managed within separate repositories that are disconnected and isolated from other contextual and external systems (i.e., data silos) preventing users and health practitioners from having an integrated view of the whole knowledge.

Cognitive computing eHealth systems (i.e., the emulation of human thought process in a computerized model applied to healthcare) can deal with all the amount of data available from IoTs to address specific issues (Riccucci *et al.*, 2005) which involve the combination of that data with users' medical history or environment information to produce actionable insights for a more targeted medical care. Moreover, eHealth can also take advantage of cognitive computing systems when services and applications use different vocabularies and models for the same concepts, thus solving issues related to interoperability and knowledge sharing (Fox, 2017). Within eHealth knowledge systems, the employment of shared ontologies constitutes a good strategy to deal with diversity. From a Semantic Web (SW) perspective, the term "ontology" denotes "a shared conceptualization of an explicit machine-readable specification". Ontologies provide a generic infrastructure for exchanging and integrating structured data and promote a creative reuse of the data which can help to respond to some of the key challenges that eHealth systems are nowadays facing.

Ontologies, as cognitive computing approach to better access knowledge, provides a common semantic description for eHealth domain concepts. Moreover, specific imaging ontologies can also be used utilized in cooperation with computer-vision techniques (Carbonaro, 2009; Carbonaro 2010) when data collected by IoT are images. Semantic technologies provide a promising way also for image analysis and interpretation, in fact, qualitative aspects of an image (e.g. low-level visual features) can be easily mapped to high-level concepts formally described by an ontology.

In this paper, we propose an ontology-based cognitive computing eHealth system which aims to provide semantic interoperability among heterogeneous IoT fitness devices and wellness appliances in order to facilitate data integration, sharing and analysis. The original contribution of our works lies in exploiting SW technologies to explicitly and formally describe the meaning of sensors data, and to facilitate the interoperability and data integration among eHealth systems.

The paper is organized as follows: the next section introduces the main previous works describing SW technologies used in the healthcare context and explores research efforts related to manage different IoT data, highlighting the main open-issues in the field. Section 3 describes the technological aspects of the context and shows the overall architecture of the proposed framework. Section 4 proposes eHealth ontologies used to describe domain concepts. Section 5 reviews our development process in order to design the ontology and describes its main characteristics. Section 6 introduces the mapping process by which data are semantically annotated according to our ontology. Finally, Section 7 provides some considerations on a case study.

2 Related works

One of the main challenging problems connected with the existing IoT applications is that devices are not (or little) interoperable with each other since their data are based on proprietary formats and do not use common terms or vocabulary. Moreover, promote interoperability, reusability, and resource sharing among IoT applications is more complicated if IoT solutions are designed by considering a single domain. SW technologies can be employed in IoT (SWoT) to overcome these challenges.

Recently, Patel *et al.* (2017) created SWoTSuite, which is an infrastructure that enables SWoT applications. It takes high-level specifications as input, parses them and generates code that can be deployed on IoT sensors (e.g., temperature sensors, transportation devices) at the physical layer and IoT actuators (e.g., heaters) and user interface devices (e.g., smartphones, dashboards) at the application layer.

Ruta *et al.* (2012) proposed a general framework for the SWoT evolved over the basic knowledge base model. An ontology along with a set of asserted facts build a knowledge base from which further knowledge can be inferred. By sharing the system infrastructure, several object classes, described using different ontologies, can co-exist in a physical environment. Resources belonging to the same domain will be described by means of the same ontology, while objects of different categories may refer to different ontologies. The main values from the use of semantic technologies in IoT can be derived through cross domain or horizontal applications rather than focusing on domain specific of vertical IoT application development.

CREDO (Fox, 2017) is a long-term research program on reasoning, decision-making, planning, and autonomy. Its primary goal is a theoretical foundation for high-level cognition understanding; the program has been inspired by clinical expertise and medicine has been central for validating the results. The paper highlighted the increased understanding of the need to capture and sharing the meaning of concepts in complex domains like medicine where great efforts are being made to establish standard terminologies and ontologies. It suggested improving CREDO system using knowledge representation and ontology design as are domain terminologies and ontological content.

From the above-mentioned papers, it is possible to underline some main issues: the semantic interoperability of eHealth connected objects and their data is crucial but still poorly widespread. Often, data concerning the connected objects (e.g., characteristics, state, properties) and data concerning the patient (e.g. vital signs, activity) are represented in a disjointed context resulting into vertical application development.

3 Semantic Web approach to eHealth

By offering a generic infrastructure for interchange, integration, and creative reuse of structured data, SW can help to cross some of the boundaries that Web 2.0 is facing. Currently, Web 2.0 offers just basic query possibilities like searching by keywords or tags. There has been a great deal of interest in the development of semantic-based systems to facilitate knowledge representation and extraction and content integration (Carbonaro, 2010; Henze *et al.*, 2004; Andronico *et al.*, 2003; 2004). We can consider semantic information representation as an important step towards a wider efficient manipulation and knowledge representation (Carbonaro, 2006). In the digital library community, a flat list of attribute/value pairs is often assumed to be available (Carbonaro, 2010). In the SW community, annotations are often assumed to be an ontology instance. Through the ontologies, the system will express key entities and relationships describing resources in a formal machine-processable representation.

In eHealth domain, there are massive information resources in which the knowledge formation process is associated with multiple data sources. However, the systems, grammar, structure and semantics of resources are heterogeneous. The idea behind SW approaches is using the Web to allow exposing, connecting and sharing data through dereferenceable Uniform Resource Identifiers (URIs). The goal of SW is to extend the Web by publishing various open datasets as Resource Description Framework (RDF) triples and by setting RDF links between data items from several sources. Using URIs, everything can be referred to and looked up both by people and by software agents. Using URIs, RDF language and Ontology Web Language (OWL) ontologies, SW technologies easily allow users to connect pieces of data, share information and knowledge on the web. Ontologies constitute the backbone of the SW. Ontologies are a means to express concepts of a given domain and the relationships among the concepts; they specify complex constraints on the types of resources and their properties.

Interoperability is the ability to interconnect and create communications between different systems and along with data; integration is one of the vital issues still unsettled in IoT. However, interoperability can be solved if communicating smart systems are semantically interoperable. This challenge is particular relevant in the eHealth where a multitude of diverse devices collect the same type of data but store and exchange them in many different ways so generating semantic and syntactic conflicts. Semantic Web of Things (SWoT) is a research field which aims to combine SW technologies to IoT by providing interoperability among ontologies and data (Jara *et al.* 2014; Pfisterer *et al.*, 2011).

Semantic data annotation is the key step for every SW project. Our framework aims to semantically annotate heterogeneous IoT fitness and life logging data collected by wearable devices and wellness appliances in order to make it integrated and machine-understandable.

Figure 1 shows the overall architecture of the proposed framework. The two core components of the entire system are the IoT Fitness Ontology (IFO) and the mapping process (i.e., the RML processor and the mapping specifications).

The primary role of the IFO ontology is to provide a formal representation of the main concepts within the IoT fitness domain. The RML processor, supplied with the mapping specifications for the various sources, consumes the IoT raw data and transforms it into an RDF graph, which is the same input data semantically annotated according to the IFO ontology.

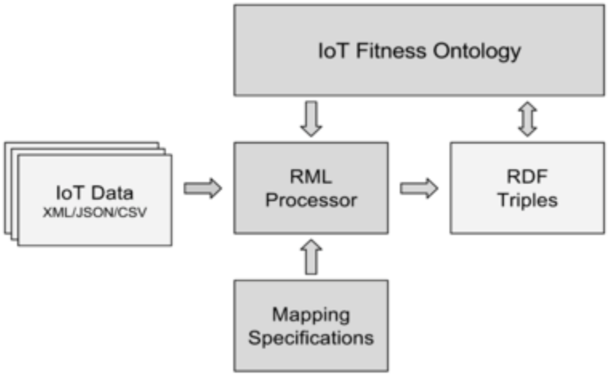

Fig. 1 - Proposed system architecture: how to build ontology using mapping and IoT raw data.

4 eHealth ontologies

Due to the extreme complexity of eHealth terminology systems, ontologies play a central role for the representation, management, and sharing of knowledge and data. In the past years, a plethora of eHealth domain ontologies has been created. Such representations are used to systematically denote, categorize, and relate eHealth concepts, allowing fitting handling of the data [46]. The number of ontologies in eHealth is constantly increasing; BioPortal provides access to a library of biomedical ontologies and terminologies developed in RDF, OWL, Open Biological and Biomedical Ontologies (OBO) formats (Smith, Ashburner *et al.*, 2007).

Below the main characteristics of SNOMED-CT and LOINC ontologies are briefly reviewed.

4.1 SNOMED-CT

The Systematized Nomenclature of Medicine-Clinical Term (SNOMED-CT) is considered as the main ontology for a standardized representation and automatic interpretation of clinical concepts, terms and relationships in the field of eHealth. The ontology covers most of the areas that are used in medical practice, including clinical findings, symptoms, diagnoses, pharmaceuticals, body structures, medical devices, social contexts, and so on. SNOMED-CT has hierarchy structure with a set of top level general concepts. All other concepts are subtypes of one these top concepts. Each concept is assigned a unique ConceptID and a Fully Specified Name (FSD). SNOMED-CT provides a consistent way to represent data that can enhance the interoperability between different systems.

4.2 LOINC

The Logical Observation Identifiers Names and Codes (LOINC) is a universal code system for laboratory test and other clinical observations. For each observation it provides a code, a short name, a long formal name and synonyms. The primary purpose of LOINC is to provide common codes and terminology to receive clinical observations from multiple sources, so that they can automatically fill the data in the right slots of their medical records, research, or public health systems.

5 Proposed ontology

The proposed IFO ontology is a lightweight extensible domain-specific ontology which aims to provide a formal representation of the most common concepts and their relationships within the IoT fitness and wellness devices, including mobile health applications.

This section briefly reviews our development process in order to design the IFO ontology and describes the main characteristics of the developed ontology.

5.1 Development process

For the development process of the IFO ontology, we followed the well-known methodology proposed by Noy and McGuiness which is a quick but complete approach for building ontologies (Noy & McGuiness, 2001). We wrote the IFO ontology in OWL, which is a W3C recommendation and the de facto standard language for publishing and sharing ontologies in the SW and we validated it using the ontology reasoner HermiT (Shearer *et al.*, 2008) to check for inconsistencies.

To identify the concepts described in the IFO ontology we considered and carefully analysed the characteristics and functionalities provided by several IoT wearable devices and wellness appliances as well as health mobile applications available in the market. The list of products and vendors that were taken in consideration during the design process includes: Apple HealthKit, Microsoft HealthVault, Google Fit, Fitbit, Jawbone, Strava, Runtastic, iHealth and Nokia Health. The result is a harmonised ontology of the most important common concepts in the domain considered. The first version of the IFO ontology consists of 93 classes, 16 object properties, 7 data properties, and 47 individuals.

5.2 Ontology Structure

The root concept of the IFO ontology is modelled by the class Episode, representing the set of the all possible events that can be measured by the IoT fitness and wellness systems. To each episode, is associated by means of OWL properties, a time reference (i.e., start time and end time of the event) and a numeric measurement value with the unit. These information are essential because they allow us to numerical quantify the object of the episode and give it a temporal collocation and duration. An example of episode could be a running activity or a treadmill session at the gym as well as a vital sign measurement such as the heart rate or the blood pressure.

The IFO ontology organizes the episodes in a hierarchical structure based on single inheritance. We consider physical activities and body measurements as important episode categories. Physical activities comprehend any kind of activity involving body movement such as running, swimming or steps taken. Body measurements, on the other hand, are relative to the physiological parameters of a person such as the body weight or the heart rate. Other categories that our ontology defines concern the sleep and the meditation. Furthermore, the IFO ontology describes complementary concepts to specify additional information regarding a single episode such as geospatial locations (i.e., acquired via GPS sensor), device characteristics, personal annotations, temporal relationships with respect to other person's life activities (e.g., temporal relationship of an episode with respect to a meal), physical activity intensity and person's information such as the gender or the date of birth.

To achieve a better integration with other systems and better specify the meaning of each class, we referred to other standardized ontologies such as SNOMED-CT. Personal information (e.g., date of birth) are based on FOAF ontology and the Basic geo (WGS84 lat/long) vocabulary was used for the geospatial locations. In order to keep the ontology simple, we avoided to import other OWL top level ontologies for concepts related to the measurements (e.g., units of measure) or for the time information (e.g., time intervals). Figure 2 shows the IFO ontology classes hierarchy as can be seen within the Protegè editor.

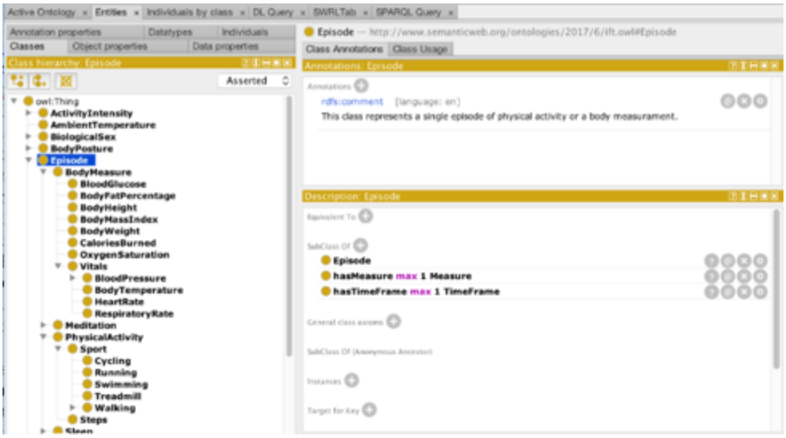

Fig. 2 - Developed ontology structure in Protegè editor.

6 Mapping

Mapping is the process by which values within the data sources are semantically annotated according to an ontology (Amardeilh, 2008). We use RML, a language to specify mappings for heterogeneous and hierarchical serializations into RDF, according to an RDF schema or ontology of the user's choice. This solution leads to mapping on the semantic level, provides a solution applicable to a broader domain and can be extended to cover a great number of file serialization. The mapping process constitutes the second core component of our framework. Essentially, a map processor consumes the input sources along with the mapping specifications to produce the RDF graph. It is important that the semantic annotation process adheres to a common standard to guarantee interoperability between different systems. RDF, the cornerstone of the SW provides a standardized means for adding metadata annotations to resources (Manola *et al.*, 2004), in order to reach semantic interoperability and integration and querying of data having heterogeneous formats. These heterogeneous data can be following leveraged in Linked Data applications. Therefore, generating RDF triples from sources having various formats is a key step for our system.

This section presents a brief overview of how data can be retrieved and the most common data serializations used by the IoT fitness devices, and then it discusses the technologies we used to implement the mapping process.

6.1 IoT data

Raw data collected by IoT devices can be manually retrieve when systems are provided with data export functionalities. IoT raw data can normally be exported in eXtensible Markup Language (XML) or Comma Separated Value (CSV) serialization formats. On the other hand, when a data export function is not directly available within the device or the mobile application, data collected by IoT systems can be downloaded from the Cloud, usually in JavaScript Object Notation (JSON) format, through RESTful APIs provided by the device vendor.

6.2 Mapping process

From a data perspective, the context of IoT, is characterized by a high heterogeneity of data representation and serialization formats. Among different vendors of IoT fitness devices the same concepts are represented using different types and stored in different formats. To address this issue, we chose to implement our mapping system using the RDF Mapping language (RML) along with the RML Processor.

RML is a mapping specification language based on RDF syntax; it derives from the R2RML language which is the W3C standard for mapping relational databases into RDF. RML is specifically designed for dealing with heterogeneous data sources. To refer to specific values within the input data sources, RML relies on a target expression language relevant to the source format. For instance, given a data source serialized in XML format, XPath can be used as an expression target language to extract the specified values, in a similar way values in a JSON source can be referenced using JSONPath.

We defined the mapping specifications for three IoT systems among the ones we used to construct the ontology (i.e., Fitbit, Apple Health and Nokia Health). In particular, the mapping rules are relative to some shared concepts among these systems (e.g., the heart rate). As an evidence of the flexibility of the mapping language, we selected the IoTs devices aforementioned because they use different formats to store the data collected. Even though we mapped only a limited number of devices, mapping definitions can be easily reused across different sources that provide similar information. As a mapping process executor, we opted for RML Mapper which is a Java implementation of an RML mapping processor. RDF Mapper already supports XML, JSON and CVS data formats, and therefore we did not need to extend or modify the existing software.

Conclusions

The cognitive computing approach proposed in this paper, allows establishing a common semantic for addressing the eHealth concepts. Ontology representation provides semantic interoperability among heterogeneous IoT fitness and life log data collected by wearable fitness devices and wellness appliances and facilitates data integration and sharing. The two core components of the framework are the IoT Fitness Ontology and the mapping system based on the RML language. SW technologies have been used also to enable advanced analytics over the IoT data. The reasoning tests performed using OWL and SWRL allow the automatic classification of classes using description logic expressions. The inferences were revised by an expert, who confirmed that they were valid based on his analysis of the available information. We are carrying out some more tests to verify that our model can efficiently implement the classification process.

The system addresses the issue of the dimension and heterogeneity in source and format of data captured by eHealth IoT devices by representing the semantics of both connected objects, the domain and their relationship to each other. Two are the main contributions of this work. First, we propose a semantic-based approach that starts with data collection, followed by knowledge extraction and semantic modelling of this knowledge, in order to explicitly describe the meaning of the sensors data. Second, we describe an eHealth system that integrates ontologies to facilitate the interoperability and data integration among different devices by illustrating the effectiveness of the proposed approach for ontology building and evaluation.

REFERENCES

Amardeilh F. (2008), *Semantic annotation and ontology population*. Semantic Web Engineering in the Knowledge Society, 424.

Andronico, A., Carbonaro, A., Colazzo, L., & Molinari, A. (2004), *Personalisation services for learning management systems in mobile settings*. International Journal of Continuing Engineering Education and Life Long Learning 14.4-5: 353-369.

Andronico, A., Carbonaro, A., Colazzo, L., Molinari, A., Ronchetti, M., & Trifonova, A. (2003), *Designing models and services for learning management systems in mobile settings*. Workshop on Mobile and Ubiquitous Information Access. Springer Berlin Heidelberg.

Carbonaro A. (2006), *Defining personalized learning views of relevant learning objects in a collaborative bookmark management system*, In Z. Ma (Ed.), Web-based Intelligent ELearning Systems: Technologies and Applications (pp. 139-155). Hershey, PA: Information Science Publishing.

Carbonaro A. (2007), *Personalized information retrieval in a semantic-based learning environment*, Social Information Retrieval Systems: Emerging Technologies and Applications for Searching the Web Effectively, Pages 270-288

Carbonaro A. (2010), *WordNet-based summarization to enhance learning interaction tutoring*, Journal of E-Learning and Knowledge Society, Volume 6, Issue 2, Pages 67-74

Carbonaro A. (2009), *Collaborative and semantic information retrieval for technology-enhanced learning*, CEUR Workshop Proceedings, Volume 535, 2009, 8p

Carbonaro A. (2010), *Improving web search and navigation using summarization process*, Communications in Computer and Information Science Volume 111 CCIS, Issue PART 1, Pages 131-138

Carbonaro, A. (2010), *Towards an automatic forum summarization to support tutoring.* Technology Enhanced Learning. Quality of Teaching and Educational Reform. Springer Berlin Heidelberg, 141-147.

Fox J. (2017), *Cognitive systems at the point of care: The CREDO program*, Journal of Biomedical Informatics.

Henze N., Dolog P. & Nejdl W. (2004), *Reasoning and Ontologies for Personalized E-Learning in the Semantic Web*, Educational Technology & Society, 7 (4), 82-97.

Islam S. R., D. Kwak, M. H. Kabir, M. Hossain, & K.-S. Kwak (2015), *The internet of things for health care: a comprehensive survey*. IEEE Access, 3:678–708.

Jara A. J., A. C. Olivieri, Y. Bocchi, M. Jung, W. Kastner, & A. F. Skarmeta (2014), *Semantic web of things: an analysis of the application semantics for the iot moving towards the iot convergence*. International Journal of Web and Grid Services, 10(2-3): 244–272.

Kim J., JW Lee (2014), *OpenIoT: An open service framework for the Internet of Things*, IEEE World Forum on Internet of Things.

Manola F., E. Miller, B. McBride (2004), *Rdf primer. W3C recommendation*, 10(1-107):6.

Mendoza, J.A., Baker, K.S., Moreno, M.A., Whitlock, K., Abbey-Lambertz, M., Waite, A., Colburn, T. & Chow, E.J. (2017), *Fitbit and Facebook mHealth intervention for promoting physical activity among adolescent and young adult childhood cancer survivors: A pilot study*. Pediatric Blood & Cancer.

Noy, Natalya F., & Deborah L. McGuinness. (2001), *Ontology Development 101: A Guide to Creating Your First Ontology*. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05.

Patel, A Gyrard, SK Datta & MI Ali (2017), *SWoTSuite: A Toolkit for Prototyping End-to-End Semantic Web of Things Applications*, Proceedings of the 26th International Conference on World Wide Web Companion, pp. 263-267.

Pfisterer, D., Romer, K., Bimschas, D., Kleine, O., Mietz, R., Truong, C., Hasemann, H., Kröller, A., Pagel, M., Hauswirth, M. and Karnstedt, M. (2011), *Spitfire: toward a semantic web of things*. IEEE Communications Magazine, 49(11):40–48.

Riccucci, S., Carbonaro, A. & Casadei, G. (2005), *An architecture for knowledge management in intelligent tutoring system*, IADIS International Conference on Cognition and Exploratory Learning in Digital Age, CELDA, 473-476.

Ruta, M., Scioscia, F., & Di Sciascio, E. (2012), *Enabling the semantic web of things: framework and architecture*. In: 2012 IEEE Sixth International Conference on Semantic Computing. IEEE.

Shearer R., B. Motik, & I. Horrocks (2008), Hermit: A highly-efficient owl reasoner. In OWLED, volume 432, 91.

Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L.J., Eilbeck, K., Ireland, A., Mungall, C.J. & Leontis, N. (2007), *The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration*, Nature Biotechnology 25, 1251 – 1255.

Je-LKS
Journal of e-Learning and Knowledge Society
The Italian e-Learning Association Journal
PEER REVIEWED PAPERS
NEW TRENDS, CHALLENGES AND PERSPECTIVES ON HEALTHCARE COGNITIVE COMPUTING: FROM INFORMATION EXTRACTION TO HEALTHCARE ANALYTICS
Vol. 14, n.1, 2018
ISSN: 1826-6223 | eISSN: 1971-8829

ADOPTING COGNITIVE COMPUTING SOLUTIONS IN HEALTHCARE

Mauro Coccoli[1]
Paolo Maresca[2]

[1] University of Genoa, Italy - mauro.coccoli@unige.it
[2] Federico II University of Naples, Italy - paolo.maresca@unina.it

Keywords: Cognitive Computing, E-Health, Healthcare Big Data, Artificial Intelligence

This paper discusses possible motivations to adopt cognitive computing-based solutions in the field of healthcare and surveys some recent experiences. From a very practical point of view, the use of cognitive computing techniques can provide machines with human-like reasoning capabilities, thus allowing them to face heavy uncertainties and to cope with problems whose solution may require computing intensive tasks. Moreover, empowered by reliable networking infrastructures and cloud environments, cognitive computing enables effective machine-learning techniques, resulting in the ability to find solutions on the basis of past experience, taking advantage from both errors and successful findings. Owing to these special features, it is perceptible that healthcare can greatly benefit from such a powerful technology. In fact, clinical diagnoses are frequently based on statistics and significant research advancements were accomplished through the recursive analysis of huge quantity of unstructured data such as in the

for citations:
Coccoli M., Maresca P. (2018), *Adopting Cognitive Computing Solutions in Healthcare,* Journal of e-Learning and Knowledge Society, v.14, n.1, 57-69. ISSN: 1826-6223, e-ISSN:1971-8829
DOI: 10.20368/1971-8829/1451

case of X-ray images or computerized axial tomography scans. As another example, let us consider the problem of DNA sequence classification with the uncountable combinations that derive from such a complex structure.

1 Introduction

Pushed by the fast and unstoppable innovation in Information and Communication Technology (ICT), we are experiencing daily evolutions in applications and services that we commonly use. This is due to both the wide availability of computational resources, and the large amount of data exchanged at high speed between a variety of heterogeneous devices and systems. Such an overwhelming progress impacts on a vast class of applications and gives way to the raise of a new wave of advanced services. In particular, the technological framework enabling this new wave can be depicted by the following keywords: Cloud, Semantic Web, Big Data, and Cognitive Computing. In this paper, we focus on cognitive computing that, in turn, relies on the cloud infrastructure and profitably implements semantic Web techniques to analyse big data, making them meaningful and transforming them in valuable information. In fact, cognitive computing systems owe their success to the capability of fast-processing huge amounts of data through the novel and sophisticated machine learning algorithms they are based on. In the present situation, healthcare is one of the pioneer fields in which cognitive computing is being applied extensively. We will present an overview of challenging research topics and we will showcase some of the results already achieved.

The remainder of the paper is organized as follows. In Section 2 a general introduction to cognitive computing is presented. Section 3 covers issues on programming such systems, while Section 4 takes into account recent cognitive computing applications in the specific field of healthcare. Conclusions follow in Section 5.

2 Cognitive Computing

From a practical point of view, we regard cognitive computing as the revamp of precedent well-funded theories that hardly found practical applications at the time of their formulation, due to the lack of computing power. This is the case, for example, of Artificial Intelligence (AI) and neural networks, which are characterized by high complexity and by the need of executing huge numbers of parallel operations in strict time frames. Traditional AI techniques rely on the model of expert systems and exploit statistics and complex mathematical model, thus require that a significant amount of operations per second is executed on large dimension data sets for their training. In this respect, cognitive computing can be considered as the revenge of AI since, nowadays, we have computing architectures suited to face large dimensional problems on large data sets. This continuous number crunching allows performing analytics and deriving new knowledge, which results in the ability to anticipate solutions in a heterogeneous class of problems. As an example, let us consider the Internet search engines. They employ such a kind of AI, to the aim of giving back information that is relevant to the users, based both on individuals' data and on the application of patterns. This involves issues such as, e.g., language contextualization, classification, clustering, entity extraction, and more. As another example, let us consider the popular sites of electronic commerce and the recommendation systems they adopt. When one is seeking an item, or immediately after a purchase, the algorithm driving the recommendation system exploits cognitive computing techniques to provide focused shopping advices related to the products one is browsing, has marked as favourite, has added to his/her wish list, also relying on users' preferences and on their purchase history (Pazzani & Billsus, 2007).

To conclude, we observe that cognitive computing systems should be regarded as a "more human" AI. In fact, they mimic human reasoning methodologies, showing special abilities in dealing with uncertainties and in solving problems that typically entail computationally heavy processes. Moreover, they expose learning capabilities, thus their knowledge base is continuously-growing and, accordingly, their reasoning ability is continuously-enhancing. Besides, cognitive computing plays an important role in improving both man-to-machine and machine-to-machine communications, and in fostering the development of new human-computer interactions models, based on the Natural User Interface (NUI) paradigm such as, e.g., conversional systems (Nishida *et al.*, 2014). Moreover, through the effective and reliable simulation of the reasoning processes, machines can be trained to learn from experts' behavior in approaching problems and from their problem-solving techniques, to become, in turn, able to train newbies and to teach humans new concepts and/or new procedures. Such intelligent systems could be used, e.g., in training and customization, or other activities that requires data analysis in order to improve both processes and products (Earley, 2015).

Furthermore, cognitive computing is thought to be the corner-stone for the future enlargement of the Internet of Things (IoT) scope (Zhang *et al.*, 2012) and, consequently, on the relevant interconnect technologies (Orii *et al.*, 2016). In fact, the expected near-future scenario is forecasting people and things to interact naturally (Coccoli & Torre, 2014), striving spoken language, while producing and consuming data performing their actions. Therefore, we need advanced analytics to gather information and extract data and vice versa, to the aim of realizing novel somehow-intelligent systems that are able to react in real time to unpredictable external stimuli with unknown origins. In this respect, we observe that the more important benefits that can derive from cognitive computing will not reside in the cognitive systems themselves, but in the coupling of cognitive systems with the surrounding environment. Then a novel era for engineering will rise, in which design will be driven by desired behaviours rather than by design constraints: *what* will make the machines rather than *how* they will be made (Holtel, 2014).

In conclusion, owing to the prospect resources and capabilities available within a working environment based on the use of cognitive computing, we can envisage the inception of a new generation of semi-autonomous systems committed to improve the quality of life, addressing critical societal issues (Mohideen & Evans, 2015), helping people in facing a variety of small complications as well as awkward jobs. Such new systems will be mainly based on the imitation of human attitudes and reasoning (Mohda *et al.*, 2011). Systems based on cognitive computing can successfully accomplish difficult tasks such as, e.g., classification, natural language processing, and data mining, thus are able to perform advanced activities such as, e.g., sentiment analysis, relationships extraction from unstructured corpus, image recognition, speech-to-text and text-to-speech conversions. Another prospected advantage is the growing confidence in humans that machines can provide reliable answers, i.e., within an acceptable range of trust, in delicate areas, such as, e.g., medicine, education or economics (Coccoli *et al.*, 2016).

3 Programming Cognitive Computing Systems

As already mentioned, cognitive computing systems owe their powerful characteristics to the recent enhancement of traditional AI pushed by the availability of new technologies, to the rapid development of new machine learning techniques, and to the large availability of data coming from a heterogeneous set of sources and devices. These reflect in the possibility of implementing effective model-based reasoning capabilities, which make cognitive systems able to perform complex tasks such as, e.g., discovery, reasoning, and multimodal understanding in a variety of domains (Banavar, 2015) such as, e.g., healthcare, insurance, and education (Coccoli *et al.*, 2017). As a consequence, cognitive computing and the relevant technology are going to play a key role in engineering systems of the future (Noor, 2015).

3.1 The Cognitive Computing Consortium

Given such premises and the forecast on the future cognitive computing development and achievements, it is necessary to set up an open working environment where researchers and IT professionals can find non-proprietary definitions that can be used as benchmark. To this aim, a cross-disciplinary group of experts from industry, academia and the analysts' community, founded the Cognitive Computing Consortium[1]. Constituents come from a mix of research centres, companies and institutions such as Synthesis and NextEra Research (founders) with Pivotal, Basis Technology, HP, IBM, BA Insight, CustomerMatrix, SAS, Interactions, Bebaio, Microsoft, and universities such as, UCSF and the Babson College. It is worthwhile noticing that most of the sponsors are companies involved in big data analysis.

3.2 Programming Cognitive Computing Systems

To the aim of spreading the adoption of cognitive computing, specific platforms and tools exist, enabling programmers to develop suited, effective and reliable systems tailored to solve their problems. In recent years, many big players in the IT scenario delivered their own cognitive computing kit and this is driving both market and technology in the direction of making such systems affordable and widely available. Among these, in the following we mention the most significant examples that are, in alphabetical order:

(i) Enterprise Cognitive Systems by Enterra. Formerly known as Cognitive Reasoning Platform (CRP), the enterprise cognitive systems framework is defined by its developers as "an artificial intelligence platform that allows organizations to capitalize on the power and potential of big data through advanced analytics and actionable insights that fundamentally inform organizations about the business, customers, and value chains in which they operate". They also claim it can easily combine "[..] the efficiency and accuracy of computational computing with the analytic and predictive abilities of human reasoning. [..] can receive extraordinary volumes of data from any source, structured and unstructured, understand the nature of the data, learn from the relationships and connections it discovers, make decisions, and take actions to achieve defined outcomes";

(ii) Deep Learning Technology Center. It is the structure owned by Microsoft where to work with the Cognitive Toolkit, which was formerly known as the Computational Networks Toolkit (CNTK), made available in open source for anyone to use in their own work on GitHub. It is depicted by its developers as "A [..] commercial-grade toolkit that trains deep learning algorithms to learn like the human brain". This tool allows creating deep learning networks for different activities and "[..] empowers you to harness the intelligence within

---

[1] http://cognitivecomputingconsortium.com

massive datasets through deep learning by providing uncompromised scaling, speed and accuracy with commercial-grade quality and compatibility with the programming languages and algorithms you already use";

(iii) DeepMind. It is the platform offered by Google that, in 2014, acquired the namesake UK-based AI company aimed to solving artificial intelligence problems. Their claim is "Solve intelligence. Use it to make the world a better place". Then, to show the effectiveness of their work, in 2015, Google announced the creation of a specific AI that learns by itself and is able to win video games. Indeed, they "were able to create a single program that taught itself how to play and win at 49 completely different Atari titles, with just raw pixels as input. And in a global first, the AlphaGo program took on the world's best player at Go - one of the most complex and intuitive games ever devised, with more positions than there are atoms in the universe - and won";

(iv) IDOL (Intelligent Data Operating Layer). It is the software layer offered by Hewlett-Packard, whose tagline is "Unified machine learning platform for enterprise search and big data analytics - text analytics, speech analytics, image analytics and video analytics". Delivered by HP, which acquired Autonomy in 2011, within their big data software platform, IDOL is offering many services and solutions for, e.g., data analysis and IoT. They claim that the "IDOL Natural Language Question Answering empowers organizations to tap into the full potential of big data by breaking down the barriers between machines and humans. It effectively unleashes the power of machine learning by enabling natural language based human-centric exchanges in delivering the contextually relevant information";

(v) Watson, by IBM. It promises to "go beyond artificial intelligence". It is a technology platform using natural language processing and machine learning to reveal insights from large amounts of unstructured data. IBM claims that Watson "can understand all form of data, interact naturally with people, and learn and reason, at scale. [..] you can analyse and interpret all of your data, including unstructured text, images, audio and video [..] you can utilize machine learning to grow the subject matter expertise in your apps and systems [..] you can provide personalized recommendations by understanding a user's personality, tone, and emotion [..] you can create chat bots that can engage in dialog".

4 Cognitive Computing Applications in Healthcare

Following the digitization process of medical records occurred in the recent years, we notice that, as other application fields, healthcare too is suffering from an explosion of information. On one hand, a huge amount of data reveals new opportunities for the advancement of research and for the effective treatment of diseases. On the other hand, such an information overload is hard to manage for both physicians and care providers. The unique reasoning ability of cognitive systems can perform detailed analysis and comparisons exploiting all the data available, thus becoming an effective companion.

As researchers, we can envisage a variety of applications in which cognitive computing can contribute to evolve healthcare but, in the current situation, its most promising feature appears to be the ability in managing huge quantity of information. In fact, cognitive system can easily overtake present solutions for big data management and decision support and these class of problems is very common in medicine, especially for prevention and diagnosis based on statistics analysis and visual pattern recognition. Most impressive and eye-catching results have been achieved in cancer diagnosis and results in this field have been proudly announced by the technology provider of the above-presented Deep Mind, Google, and Watson. As already stated, recognizing and classifying images is another peculiar ability of cognitive systems (Teo *et al.*, 2012) and this feature is a strategic asset in the prevention of cancer pathologies, especially for what concerns breast cancer, lung cancer and prostate cancer (Strickland, 2013). In fact, a cognitive system that "sees" is a valuable support, relieving the doctor from the task of analysing many hundreds of thousands of documents about the same pathology in a little time-frame as well as for providing the semantic interpretation of diagnostic images (Ogiela *et al.*, 2006). However, it is worth pointing out at this point how the application of the cognitive system in the medical field in particular finds a series of barriers and resistances by physicians and nurses, mainly due to the lack of basic computer skills. Nevertheless, there is evidence that where cognitive systems are used for healing, a 50% improvement in results is observed, while hospitalization costs are reduced by half (Gatenbein, 2014). In addition, there is also a reduction in diagnostic errors, especially in carcinogenic pathologies. This last observation is changing the focus of operators, because in the United States the first cause of death is due to diagnostic errors.

To clarify the situation, we propose a literature review of cognitive computing solutions and related technologies applied to the healthcare. We will not enter in details of specific solutions and methodologies adopted, yet we will remain at an abstraction level where the benefits are highlighted and foreseen for the future development of novel decision support systems and autonomous services as well, to the aim of improving individuals' quality of life and health. In the following, with no ambition to be exhaustive, we report some recent works that illustrate the above-cited ability. It is worthwhile noticing that many articles in the clinical literature refer to systems based on IBM Watson and we reserved a specific section for those.

4.1 Decision Support

Decision making and decision support can benefit from cognitive computing capabilities. One example of this is reported in the paper "Temporal Modeling in Clinical Artificial Intelligence, Decision-Making, and Cognitive Computing: Empirical Exploration of Practical Challenges", Bennett and Doub (2014) describe a decision system based on Markov Decision Processes and the use of neural networks, implementing the so-called temporal modelling approach, that allows capturing certain aspects of human cognition. In this respect, the computing system should resemble the same process, hence cognitive computing solution re expected to improve performances of the above-cited approach. Other methods, such as, e.g., Interactive Metric Learning (IML) are described in the paper "IBM's Health Analytics and Clinical Decision Support", by Kohn *et al.* (2014). Another core aspect in healthcare involving decision support systems is related to the management of patient records. In their paper "Cognitive computing for electronic medical records", Devarakonda and Mehta (2016) describe the problems due to the overload of information in Electronic Medical Record (EMR) systems and the inability to make sense of this information to provide the best care for their patients. They identify the solution in the use of cognitive systems designed to perform advanced analysis on the patient record data. These may also require the understanding of natural language questions about the patient record content, helping physicians to automatically identify urgent abnormalities, and provide precise causes for such abnormalities. In such a cognitive computing view, the EMR is transformed in an active entity that helps making decisions, leveraging the large amount of knowledge within the medical sciences, drug information, and medical ontologies.

4.2 Big Data and Analytics

In the editorial "Big Data and Analytics", by Tan *et al.*, the authors put in evidence how the digitization process occurred in health and in the management of patient data, as well as the rapid adoption of health information systems have led to the generation of huge volumes of primary and secondary data within the health care industry, that cannot be processed and managed by traditional data processing tools and that have to be duly managed, in order to make them a valuable asset for both improving the therapy effectiveness and advancing research to enhance prevention and health outcomes, also for reducing costs (Tan *et al.*, 2015). In this respect, they introduce health analytics as "the systematic use of health data and related business insights developed through applied analytical disciplines (e.g. statistical, contextual, quantitative, predictive, cognitive, other models) to drive fact-based decision making for planning, management, measurement and learning" (Cortada *et al.*, 2012). Consequently, health analytics require a variety of statistical techniques borrowed from modelling, machine learning, data mining, to analyse current and historical facts to make predictions about unknown events. Then, Chen *et al.*, in their paper "IBM Watson: how cognitive computing can be applied to big data challenges," extend the scope of cognitive computing to the entire life sciences field (Chen *et al.*, 2016), pondering on issues that life sciences researchers have to cope, detailing how cognitive technologies can help finding new solutions to aggregate big data for a better understanding of the latent information they may contain.

4.3 Watson in Healthcare

In the paper "Paging Dr. Watson: IBM's Watson Supercomputer Now Being Used in Healthcare", Lee (2014) outlines how the supercomputer has moved on to practical applications and why it was "taught" to understand the complexities of healthcare, putting emphasis on using the term *taught* rather than *programmed*. After a brief history of cognitive computing in history, practical uses for Watson in are duly listed, such as, cancer research, supply chain management, and consumer empowerment to help create better outcomes in healthcare. Significant case studies are presented, including the work done with the Memorial Sloan-Kettering Cancer Center (MSKCC) and with the MD Anderson Cancer Center (University of Texas), which both experienced that the large amount of data collected from their patients and stored in heterogeneous systems were essentially useless due to the inability of merging the results achieved by the oncology team with the clinical trial data. Another application of Watson in oncology is reported in the paper "Envisioning Watson as a Rapid-Learning System for oncology" (Malin, 2013), which also, emphasizes on the unprecedented reasoning ability of the cognitive computing system, using machine learning to determine how to weigh clinical factors in patients, to the aim of identifying the more suited treatment options and give a decision support to physicians. In practice, Watson was trained to do this, similar to a medical school student, which learns by observing more experienced physicians.

Besides, we observe that these successful applications are leading to future developments involving the same technology. In fact, we highlight that the New York Genome Centre (NYGC) and IBM are collaborating (Ratner, 2015) to analyse genetic data to speed up the process of treatment for patients with brain cancer (Douglass & Kearns, 2017). The IBM Watson cognitive computing system will be trained to analyse genomic data from a small group of patients with glioblastoma diagnosis, one of the most aggressive and malignant brain tumours. Its cognitive abilities will be used to analyse gene sequence variations between normal brain tumour biopsies, medical information, and clinical records to help physicians locate a variety of treatments and tailor the type of cure for specific cancer. Then, applying cognitive computing power accelerates the ability to address personalized cure for fatal diseases such as cancer. Another interesting development filed is enhancing medicine in developing countries where 70% of new cases of brain cancer occur. In India, there is only one oncologist for about 1,600 patients. A cognitive system such specifically designed for Oncology (Manipal, 2017) can help with large numbers of patients. The Manipal Hospital in India is one of the private hospital chains that treat 200,000 patients a year. A physician with Watson's help takes only 20 seconds to collect information about a patient. This is a big difference because it allows very fast to give patients a cure. In addition to the speed we have much more precision: mistakes in formulating a diagnosis are reduced. Currently, IBM Watson for Oncology is used also in China, Thailand, Finland. Important results in a different field are reported in Barrow (2017). Barrow Neurological analysed 1,500 genes by discovering that five of them had never been connected to SLA. Moreover, IBM actively collaborates with the New York Collaborative Care Centre to develop a health management platform (Douglas & Adigun, 2017, CNYCC, 2017). UNC Lineberger is another comprehensive cancer centre, which adopted cognitive solutions to accelerate DNA analysis and inform personalized treatment options for patients (UNC, 2017).

Conclusions

From this overview, we can argue that cognitive computing in healthcare is a hot and promising topic. Both academics and industry are making big efforts to improve the performances of current systems and to propose novel solutions based on the profitable exploitation of big data. However, we put in evidence that most of the reported experiences are from United States where the healthcare system is organized in a peculiar manner, which is quite different from the majority of other Countries.

Furthermore, unfortunately, there is still a lack in infrastructural settings, the availability of open big data, and in general the minimum requirements for the hardware to effectively run such systems are still high, despite they can rely on modern and sophisticated cloud-based architectures. Nevertheless, cloud computing is expected to uphold its rapid growth in the very next future so that we can forecast the wide availability of affordable services for many applications. This will be one of the main pillars to base the diffusion of cognitive systems on, and will ease the penetration of such a novel variety of systems that will foster new services and will bring disruption in many settled paradigms. The cognitive healthcare will have a strong impact on cloud evolution. In fact, there will be a need to create an optimized cloud for all cognitive data - a hybrid and secure cloud. It is worth considering that in addition to the cloud, we also need to redesign the data architecture due to the heterogeneity of medical data. This is because in medicine 90% of data is image and 80% of medical data is not available on the Web, also due to security and privacy issues. Finally, cognitive health care will have a very strong impact on industry.

REFERENCES

Banavar, G.S. (2015), *Watson and the era of cognitive computing*, Proceedings of the 2015 IEEE International Conference on Pervasive Computing and Communications (PerCom), St. Louis, MO, 95-95.

Barrow Neurological Institute (2017), *Barrow, IBM Watson health ID new genes involved in ALS*, available at the URL: https://www.barrowneuro.org/press-releases/barrow-ibm-watson-health-new-genes-als, last visited, Nov., 2017.

Bennett, C.C and Doub, T.W. (2014), *Temporal modeling in clinical artificial intelligence decision-making and cognitive computing: Empirical exploration of practical challenges*, in Proceedings of the 3rd SIAM Workshop on Data Mining for Medicine and Healthcare (DMMH).

Chen, Y., Argentinis, E. and Weber, G. (2016), *IBM Watson: how cognitive computing can be applied to big data challenges*, in Life Sciences Research, Clinical Therapeutics, 38(4), 688-701.

CNYCC, The Central New York Care Collaborative (2017), *Central New York Care Collaborative Announces Partnership with IBM Watson Health on Regional Population Health Management System*, available at the URL: https://cnycares.org/news/ibm-watson-health-regional-phm, last visited, Nov., 2017.

Coccoli, M., Maresca, P. and Stanganelli, L. (2016), *Cognitive computing in education*, in Journal of e-Learning and Knowledge Society, 12(2), 55-69.

Coccoli, M., Maresca, P. and Stanganelli, L. (2017), *The role of big data and cognitive computing in the learning process*, in Journal of Visual Languages & Computing, 38(1), 97-103.

Coccoli, M. and Torre, I. (2014), *Interacting with annotated objects in a semantic web of things application*, in Journal of Visual Languages & Computing, 25(6), 1012-1020.

Cortada, J.W., Gordon, D. and Lenihan, B. (2012), *The value of analytics in healthcare: From insights to outcomes*, IBM Global Business Services, Life Sciences and Healthcare, Executive Report.

Devarakonda M.V. and Mehta, N. (2016), *Cognitive computing for electronic medical records*, in Healthcare Information Management Systems, Part of the series Health Informatics, 555-577.

Douglass, C. and Adigun, B.J. (2017), *Central New York Care Collaborative (CNYCC) chooses IBM Watson care manager to improve health across the Region*, available at the URL: http://www-03.ibm.com/press/us/en/pressrelease/51646.wss, last visited, Nov., 2017.

Douglass, C. and Kearns, K. (2017), *IBM and New York Genome Center to create comprehensive, open cancer data repository to tap cognitive insights from Watson*, available at the URL: https://www-03.ibm.com/press/us/en/pressrelease/49232.wss, last visited, Nov., 2017.

Earley, S. (2015), *Executive roundtable series: machine learning and cognitive computing*, in IT Professional, 17(4), 56-60.

Gatenbein, R.E. (2014), *Watson, come here! The role of intelligent system in Healthcare*, in Proceedings of the World Automation Congress, 1-4.

Holtel, S. (2014), *More the end of information overflow: how IBM Watson turn upside down our view on information appliances*, Proceedings of the 2014 IEEE World Forum on Internet of Things (WF-I.T), 187-188.

Kohn, M.S., Sun, J., Knoop, S., Shabo, A., Carmeli, B., Sow, D., Syed-Mahmood, T. and Rapp, W. (2014), *IBM's health analytics and clinical decision support*, in IMIA Year book, 154-162.

Lee, H. (2014), *Paging Dr. Watson: IBM's Watson supercomputer now being used in healthcare*, in Journal of AHIMA, 85(5), 44-47.

Malin, J.L. (2013), *Envisioning Watson as a rapid-learning system for oncology*, in Journal of Oncology Practice, 9(3), 155-157.

Manipal Hospitals, (2017), *Manipal Hospitals is the first healthcare provider in India to adopt Watson for oncology for upgradation of cancer care*, available at the URL: https://www.manipalhospitals.com/blog/ibm-watson-for-oncology, last visited, Nov., 2017.

Modha, D.S., Ananthanarayanan, R., Esser, S.K., Ndirango, A., Sherbondy, A.J. and Singh, R. (2011), *Cognitive computing, Communications of the ACM*, 54(8), 62-71.

Mohideen, R. and Evans, R. (2015), *Shaping Our Technological Futures*, in IEEE Technology and Society Magazine, 34(4), 83-86.

Nishida, T., Nakazawa, A., Ohmoto, Y. and Mohammad, Y. (2014), *History of conversational system development, in Conversational Informatics: A Data-Intensive Approach with Emphasis on Nonverbal Communication*, 43-62, Springer, Tokyo.

Noor, A.K. (2015), *Potential of cognitive computing and cognitive systems*, Open Eng., 5, 75–88.

Ogiela, L., Tadeusiewicz, R. and Ogiela, M.R. (2006), *Cognitive computing in intelligent medical pattern recognition systems*, in Intelligent Control and Automation, Springer Berlin Heidelberg, 851-856.

Orii, Y., Horibe, A., Matsumoto, K., Aoki, T., Sueoka, K., Kohara, S., Okamoto, K., Yamamichi, S., Hosokawa, K. and Mori, H. (2016), *Advanced interconnect technologies in the era of cognitive computing*, Proceedings of the Pan Pacific Microelectronics Symposium (Pan Pacific), 1-6.

Pazzani, M.J. and Billsus, D. (2007), *Content-based recommendation systems*, in The Adaptive Web, P. Brusilowsky, A. Kobsa and W. Nejdl (Eds.), LNCS 4321, 325-341, Springer-Verlag, Berlin Heidelberg.

Ratner, M. (2015), *IBM's Watson group signs up genomics partners*, in Nature Biotechnology, 33(1), 10-11.

Strickland, E. (2013), *Watson goes to Med School*, IEEE Spectrum, 50(1), 42-45.

Tan, S.S.-L., Gao, G. and Koch, S. (2015), *Big data and analytics in healthcare, Methods of Information in Medicine*, 54(6), 546-547.

Teo, C.L., Yang, Y., Daumè, H., Fermuller, C. and Aloimonos, Y. (2012), *Towards a Watson that sees: language-guided action recognition for robots*, in Proceedings of the International Conference on Robotics and Automation (ICRA), 374-381.

UNC Lineberger Comprehensive Cancer Center (2017), *Partnering with IBM and Watson to accelerate DNA analysis and inform personalized treatment*, available at the URL: https://unclineberger.org/news/partnering-with-ibm-and-watson-to-accelerate-dna-analysis-and-inform-personalized-treatment, last visited, Nov., 2017.

Zhang, M., Zhao, H., Zheng, R., Wu, Q., and Wei, W. (2012), *Cognitive Internet of Things: concepts and application example*, in International Journal of Computer Science, 9(6-3), 151-158.

Je-LKS
Journal of e-Learning and Knowledge Society
The Italian e-Learning Association Journal
PEER REVIEWED PAPERS
NEW TRENDS, CHALLENGES AND PERSPECTIVES ON HEALTHCARE COGNITIVE COMPUTING: FROM INFORMATION EXTRACTION TO HEALTHCARE ANALYTICS
Vol. 14, n.1, 2018
ISSN: 1826-6223 | eISSN: 1971-8829

PRELIMINARY DATA ANALYSIS IN HEALTHCARE MULTICENTRIC DATA MINING: A PRIVACY-PRESERVING DISTRIBUTED APPROACH

Andrea Damiani[1], Carlotta Masciocchi[2], Luca Boldrini[2], Roberto Gatta[2], Nicola Dinapoli[1], Jacopo Lenkowicz[2], Giuditta Chiloiro[2], Maria Antonietta Gambacorta[2], Luca Tagliaferri[1], Rosa Autorino[1], Monica Maria Pagliara[3], Maria Antonietta Blasi[3], Johan van Soest[4], Andre Dekker[4], Vincenzo Valentini[2]

[1] Polo Scienze Oncologiche ed Ematologiche, Università Cattolica del Sacro Cuore - Fondazione Policlinico Universitario Agostino Gemelli, Roma - Italy
[2] Polo Scienze Oncologiche ed Ematologiche, Istituto di Radiologia, Università Cattolica del Sacro Cuore - Fondazione Policlinico Universitario Agostino Gemelli, Roma - Italy
[3] Polo Scienze dell'invecchiamento, neurologiche, ortopediche e della testa-collo, Istituto di Oftalmologia, Università Cattolica del Sacro Cuore, Fondazione Policlinico Universitario Agostino Gemelli, Roma - Italy
[4] Maastricht University Medical Center, Radiation Oncology MAASTRO-GROW School for Oncology and Development Biology, Maastricht, Netherlands

Keywords: distributed learning, distributed preliminary analysis, privacy-preserving, healthcare, data mining

The new era of cognitive health care systems offers a large amount of patient data that can be used to develop prediction models and clinical decision support systems. In this frame, the multi-institutional approach is strongly encouraged in order to reach more numerous samples for data mining and more reliable statistics. For these purposes, shared ontologies

for citations:
Damiani A. *et al.* (2018), *Preliminary Data Analysis in Healthcare Multicentric Data Mining: a Privacy-preserving Distributed Approach,* Journal of e-Learning and Knowledge Society, v.14, n.1, 71-81. ISSN: 1826-6223, e-ISSN:1971-8829
DOI: 10.20368/1971-8829/1454

need to be developed for data management to ensure database semantic coherence in accordance with the various centers' ethical and legal policies. Therefore, we propose a privacy-preserving distributed approach as a preliminary data analysis tool to identify possible compliance issues and heterogeneity from the agreed multi-institutional research protocol before training a clinical prediction model. This kind of preliminary analysis appeared fast and reliable and its results corresponded to those obtained using the traditional centralized approach. A real time interactive dashboard has also been presented to show analysis results and make the workflow swifter and easier.

1 Introduction

In the new era of cognitive health care systems, a massive amount of previously unavailable clinical variables is available for each patient and needs to be managed (e.g. by Electronic Health Records, clinical research, pathology reports, medical reports etc.) (Patel & Kannampallil, 2014) . These data can be successfully used to develop prediction models in order to produce decision support systems for clinicians (Lambin, *et al.,* 2016). However, even if a high number of covariates represents an opportunity to investigate new relations, it also poses new challenges, starting with the higher number of patient records needed in order to achieve an adequate level of statistical significance and to enable researchers to perform model validation (Lambin, *et al.,* 2013). A sufficient number of patient records is usually available only via multi-institutional data sharing. With this approach, datasets coming from different institutions are sent to a central repository and consolidated into a single database. In order to achieve this goal, data sharing is performed by a standardized data collection system, with a shared terminological system ensuring semantic coherence (Meldolesi, *et al.,* 2014), in a privacy-preserving environment, thus achieving both usability and safety of health data, in accordance with ethical and legal requirements by local and international regulations, as in the U.S. (Korn, 2002) and in the EU (Carey, 2009). Several techniques have been proposed, such as encryption for data anonymization (Gkoulalas-Divanis *et al.,* 2014), randomization methods or k-anonymity models and l-diversity (Aggarwal & Philip, 2008). Unfortunately, such methods reduce the granularity of representations in order to increase the privacy preservation of data (Aggarwal & Philip, 2008).

Distributed Learning (DL) techniques may be a good solution to privacy-related issues in performing data analytics through the use of multi-institutional big data: they preserve patients' privacy and data ownership in training prediction models by leaving all data within the originating institutions. This approach, under some conditions, obtains the same results as the classical centralized approach (Deist, *et al.,* 2017). Boyd *et al.* (Boyd *et al.*, 2011) developed a significant class of algorithms implementing a distributed method for the support vector machine, LASSO and logistic regression using the Alternating Direction Method of Multipliers (ADMM). In practice, several applications on clinical data analytics (Lu *et al.*, 2015; Jochems *et al.*, 2016; Deist *et al.*, 2017; Damiani *et al.,* 2015) have recently been published using this approach; the architecture typically involves two components: site and master. In these applications, patient data were stored at each site and only cumulative statistics were exchanged with the central server (the master). The master then computed the new parameters, which were sent to the single site, and tuned the computation until a convergence criterion was reached (figure 1). This is an example of server-client architectures (Dai, *et al.*, 2018).

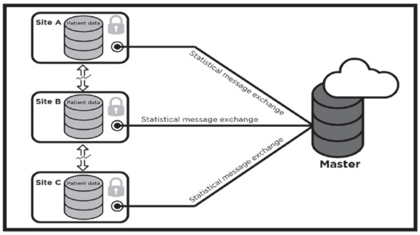

Fig. 1: Statistical message exchanges among sites and master. Patient data are stored at each site and only statistical messages are exchanged between each site and master

The large majority of publications in the DL field focus mainly on learning algorithm development and usually omit all aspects relative to preliminary data analysis. Especially in the health care field, preliminary data analysis is an essential step before training prediction models.

In any multicentric research effort in the field of health care, an initial analysis phase on enrolled patients, in which the researcher inspects available data in order to detect abnormal behaviors, contradictory trends of a given covariate across sites, or peculiar correlations of pairs of covariates (e.g. two continuous covariates exhibit a positive correlation at site A while simultaneously showing a negative correlation at site B) is necessary. These abnormalities, if left undetected, could lead to a reduced quality of the predictive model learned by the main algorithm and ultimately undermine the value of the whole research effort. In a distributed setting, such need is amplified as the researcher can only have direct access to his/her own local dataset. This means that a preliminary analysis step with a distributed approach is very important in the development of data value awareness and data quality enhancement.

1.1 Data issue

The majority of the prediction models proposed in clinical literature are developed on well-defined cohorts of patients with specific protocol designs (Steyerberg, 2008). In the case of traditional multi-institutional studies, the protocol of the study defines the policies by enrolling a subgroup of patients and is shared among the centers in order to ensure a homogeneous set of clinical cases is recruited. The level of homogeneity is generally verified during the descriptive statistics phase through the application of statistical tests such as Chi-Square, T-test or Mann Whitney test. The choice of the most appropriate statistical test depends on several factors, such as sample number or the type (numerical, ordinal, categorical, etc.) of covariate to be analyzed. This preliminary distribution verification is essential to detect possible bias in patient recruitment by a specific center. Furthermore, it ensures the reliability and reproducibility of the model across new samples coming from the same predefined target population. Some possible aims of investigation for this step are: are one or more covariates distributed in the same way across sites? Shouldn't these two covariates show the same kind of mutual correlation across sites?

The distributed privacy-preserving version of two tools adopted in descriptive statistics are proposed: a distributed version of Chi-Square test and an investigation tool using linear or logistic regression models across sites. These algorithms have been tested on real clinical data in order to analyze covariate distributions and correlations across sites, verifying the homogeneity of the data of a selected protocol and enhancing their quality. During the experiment, these tools were used by an expert in order to detect data anomalies.

2 Material and Methods

2.1Chi-Square test and linear and logistic regression model

The Chi-Square test is a statistical test which compares the distribution between two binned numeric or categorical variables. Its purpose is to accept or reject a null hypothesis (Ho). This assumption is quantified by a p-value statistic parameter. If the Ho is rejected ($p<0.05$) a statistically significant difference between two covariate distributions is observed. In the traditional approach and in case of multi-institutional data sharing, datasets are sent to a central repository and are centralized into a unique big dataset. In this case, where the data are accessible, each numeric covariate is divided into predefined intervals. In case of categorical or binary variables the interval corresponds to the variable categories. Accordingly with each binning, the number of occurrences in each interval value is calculated. Considering two binned datasets, let S_i be the number of occurrences in bin *i-th* for the first dataset and D_i the number of occurrences in bin *i-th* for the second dataset, the Chi-Square statistics is:

$$X^2=\sum_{i=1}^{k}\frac{\left(S_i*\sqrt{\frac{D}{S}}-D_i*\sqrt{\frac{S}{D}}\right)^2}{\frac{S}{D}} \quad (1)$$

where

$$S=\sum_{i=1}^{k}S_i \qquad D=\sum_{i=1}^{k}D_i \quad (2)$$

And the degrees of freedom (*dof*):

$$dof=(NumberofDataset-1)*(TotalBin-1) \quad (3)$$

Given the (1) and *dof* (3) values, the Chi-Square table distribution or its relative function available in the statistical analysis tool (such as the *pchisq* function available in "R" statistical software) can be used to evaluate the p-value statistic parameter.

2.2 Data access and distributed infrastructure

The three simulated sites collected the clinical data using an in-house software called BOA (Tagliaferri *et al.*, 2016). The aim of this software is an ontology-based standardized data collection able to improve data quality and allow cooperation among different institutions.

To this purpose, data was stored in a PostgreSQL database (version 9.4.1) and a learning connector was used for each simulated sites (Varian Medical Systems, Palo Alto, USA). Registry data (e.g. name and surname) were stored locally into a different database after a de-identification of the Patient's ID in order to increase privacy preservation of data. Through the learning connector, each site then queried the local data using SQL language and exchanged messages with the master.

The connection between each learning connector and the general server was guaranteed by a server-client architecture called Varian Learning Portal (VLP), developed by Varian Medical System company, through which intermediate statistic results were asynchronously exchanged among the master and sites. The researchers interact with the VLP using a web-based interface[1] in which they can upload their distributed algorithms and run simulations. The VLP automatically transmits the single site's algorithm to each other site and the master's to the cloud service. The site algorithm communicates with the learning connector and the master's algorithm, which runs on the VLP, can exchange intermediate statistic results, backwards and forwards with each single simulated site.

[1] (https://www.varianlearningportal.com/VarianLearningPortal/)

2.3 Distributed Chi-Square test and linear and logistic models implementation

As mentioned in section 2.1, the traditional Chi-Square test requires patient data to be accessible. In this section, we propose a distributed Chi-Square test in which patient data never leave the single originating site and health data security is assured. This approach has been used in order to analyze the distribution of the same covariates across each combination of sites.

Each site calculates the occurrences S_i and D_i by accessing its local database and without sharing patient data at each iteration. The statistical parameters of the test (e.g. Chi-Square statistics and p-values) are calculated on the master's side and a result, identical to that obtained using a centralized approach, is generated. Supposing that "M" sites are given, the occurrences S_i and D_i are calculated for each combination of two sites and sent back to the master. The master will then aggregate the statistics received from the sites, calculating the results of the Chi-Square statistic based on equation 1, 2 and 3 and the corresponding p-values. A value lower than 0.05 is considered statistically significant. Details of the proposed distributed Chi-Square test for numeric covariates iteration by iteration are listed in table 1. Regarding binary covariates, only iteration 3 and iteration 4 are used. Correlations between couples of covariates were evaluated either with local logistic or linear models, according to covariate types: linear (for both the numeric covariates) or logistic (for numeric and/or binary covariates) regression models were trained at each site. The beta coefficients and p-value parameters were then sent back to the master. These methods, in addition to distributed 2, could help researchers identify cohort differences among sites.

The proposed code was entirely developed using R version 3.3.1. The results were visualized using a dashboard called Web-based dIstributed statistics REsults (WIRE). WIRE allows the interactive visualization of the distributed descriptive statistic results in real time, offering graphical tools (see figure 2 as an example). The dashboard consists of two parts: site distributed and master distributed statistic results. In the first section, the distributed base statistic results for each site are reported in terms of the number of patients and covariate ranges. In the second section, the distributed base statistic results for the master are visualized in terms of the total number of patients, cumulative covariate ranges, distributed Chi-Square test and local linear and logistic regression models.

WIRE is supported by several browsers: Google Chrome, Mozilla Firefox, Safari and Internet Explorer and its design and development were implemented using the "Shiny" R package which develops interactive web applications simply.

Table 1
MESSAGE EXCHANGES AMONG MASTER AND SITES: ITERATION BY ITERATION

	Message exchanges among master and sites: iteration by iteration
Iteration 1	Each site calculates covariate ranges (in terms of minimum and maximum value for each covariate) and number of patients, then sends the intermediate statistic results to the master.
Iteration 2	The master receives intermediate statistics from each site. Considering each combination of two sites each time (e.g. site A and site B), the master aggregates and sends back to each site 3 values for each covariate: the maximum absolute value, the minimum absolute value and the number of bin calculated as $numBin = \text{int}\left(1 + max\left(\log\left(N_a\right), \log\left(N_b\right)\right)\right)$ where N_a and N_b are the number of patients of site A and site B respectively.
Iteration 3	Each site calculates the predefined interval values by creating a normal distribution using the number of bin, maximum and minimum values received from the master for each covariate and each combination. The number of occurrences S_i and D_i are evaluated for each value of the interval. These values are finally sent back to the master.
Iteration 4	The master receives the occurrences from all sites. For each covariate and each combination the Chi-Square statistics (based on equation 1), S, D (based on equation 2) and dof parameters (based on equation 3) are evaluated. The final p-values are then calculated using R statistical software.

3 Case study

Clinical standardized data from 234 uveal melanoma patients treated with brachytherapy were used for the purpose of our investigation. The inclusion criteria were: dome-shaped melanoma, distance to the Fovea>1.5 mm, tumor thickness>2 mm and follow-up>4 months. Three variables were used in this experience: the presence of diabetes (binary variable: yes versus no), the tumor volume (numeric variable) and the tumor distance to the fovea (numeric variable). The collected dataset was then randomly split into three databases to simulate three different sites (Site A, Site B and Site C). 83 patients were assigned to Site A, 119 to Site B and 32 to Site C. Each dataset was then archived on an independent workstation with a proper learning connector installed to simulate the existence of 3 different institutions and the learning environment described in section 2.2 was recreated. The primary aim of our experiment was to simulate the event in which 3 centers are developing a common predictive model using this infrastructure. The algorithms reported in section 2.3 were applied in order to test the distributed databases' homogeneity before the application of a distributed predictive model. The distributed Chi-Square test (see section 2.3) was applied for each covariate and for each pair of sites (e.g. combination site A-site B; site A-site B and combination site B- site C). The results of such analysis, visualized through the WIRE interface, showed some heterogeneity in terms of distribution of the "volume", "distance to fovea" and "diabetes" covariates as shown in table 2. P-values were calculated for the different combination of sites and appeared to be lower than 0.01 for the "volume" covariate in the combination of sites A-C B-C, lower than 0.01 for the "distance to fovea" covariate for the combination for sites A-C B-C and for the "diabetes" for all combinations analyzed.

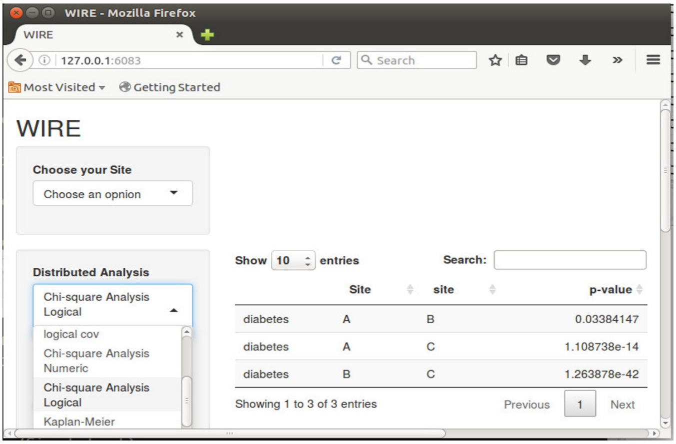

Fig. 2 - Snapshot of WIRE graphical tools. Users can visualize site and master distributed statistic results in real-time in order to check covariate distributions.

Table 2

$\chi 2$ P-VALUE DISTRIBUTED ANALYSIS FOR EACH COVARIATE AND FOR EACH COMBINATION OF SITES ARE REPORTED. THE STATISTICAL SIGNIFICANT VALUES ARE REPORTED AS BOLD VALUES

Combination sites	$\chi 2$ p-value		
	Volume	Distance to Fovea	Diabetes
A-B	0.07	0.2	**0.03**
A-C	**< 0.01**	**< 0.01**	**< 0.01**
B-C	**< 0.01**	**< 0.01**	**< 0.01**

The local linear and logistic regression model was then trained. A statistically linear correlation between "volume" and "distance to fovea" was also observed and more specifically, site B and C showed an inverse correlation when compared to site A (negative slope parameter). These results may or may not represent an alarming signal about the data and suggest that further investigation is needed before proceeding with the use of the combined data. Having observed these differences, we decided to start the aforementioned iterative data processing in order to identify and solve the heterogeneity causes, allowing the researchers to employ previously unusable data. Furthermore, thanks to the help of an expert not involved in the patient enrollment phase who checked the descriptive statistics of single centers' covariates on the WIRE interface (and therefore without accessing the single databases), we realized that the sites for numerical covariates ("volume" and "distance to fovea") had not fully respected the shared enrollment protocol, including some non-eligible patients. Having removed those patients from the database, we re-ran the tests on a total of 197 patients (site A:68; site B:65; site C:64), obtaining three homogeneous datasets in which no differences in terms of distributions among the single covariates or covariate correlation across the sites were found. Using this last patient subset, we compared the Chi-Square statistics and the observed p-value results through the distributed and centralized approaches. The results are reported in table 3. The difference between the distributed and centralized p-value was less than 10^{-16}. Finally, the response time for the distributed preliminary analysis tool appeared to be very short ($t < 0.05$ s), allowing to support a real time dataset investigation.

Table 3
THE CENTRALIZED AND DECENTRALIZED ALGORITHMS WERE IMPLEMENTED USING R. BOTH EXPERIMENTS WERE PERFORMED ON THE SAME DATASET. RESULTS SHOW THAT BOTH MODELS RESULTED IN IDENTICAL CHI-SQUARE STATISTICS AND P-VALUE COEFFICIENTS.

Features	Distributed chi-square		Centralized chi-square		Combination
	$\chi 2$	p-value	$\chi 2$	p-value	
Volume	1.783	0.878	1.783	0.8782	A-B
Distance to Fovea	5.363	0.373	5.363	0.3731	
Diabetes	0.079	0.778	0.007	0.7784	
Volume	2.741	0.739	2.741	0.739	A-C
Distance to Fovea	3.026	0.695	3.026	0.695	
Diabetes	1.540	0.214	1.540	0.214	
Volume	9.869	0.007	9.869	0.079	B-C
Distance to Fovea	5.694	0.337	5.694	0.337	
Diabetes	1.131	0.287	1.131	0.287	

4 Discussion

The proposed study addresses just one of the potential applications of distributed preliminary analysis on data before training a distributed prediction model. A mathematical approach that applies a strict privacy- preserving policy has been proposed which makes those privacy preservation barriers less difficult to manage. The results were visualized using the WIRE dashboard. It was successfully used by an expert to detect discrepancies compared to the agreed-upon research protocol. The very short running time and the achievement of the same results when compared to the centralized approach suggest that the application of this solution is workable. This approach will greatly facilitate the collaboration among institutions characterized by different ethical, legal requirements and policies on clinical data management. Some limitations of this approach are: the installation of statistical software R is necessary to run the algorithms and to compile the WIRE interface, which does not allow for running the algorithms manually. These are launched using command line codes and involvement of information technologists is therefore required. The three different sites were only simulated.

Conclusion

DL techniques may be a good solution to privacy-related issues in performing data analytics through the use of multi-institutional big data. Chi-Square test and integrating logistic and regression models were proposed as a necessary step in order to detect data heterogeneity. The technology discussed in this paper allowed clinicians to detect major abnormalities in the covariate distributions across sites, just by looking at the dashboard and without actually accessing the data. In future works, the application of these methods with model development by using real distributed sites will be mandatory.

REFERENCES

Aggarwal C.C., & Philip S.Y.(2008), *A general survey of privacy-preserving data mining models and algorithms*, In Privacy-preserving data mining(pp. 11–52), Boston, Springer.

Boyd S., Parikh N., Chu E., Peleato B., et al.(2011), *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Foundations and Trends in Machine Learning, 3(1), 1–122.

Carey P.(2009), *Data protection: a practical guide to uk and eu law*, Oxford, Oxford University Press.

Dai W., Wang S., Xiong H., & Jiang X.(2018), P*rivacy preserving federated big data analysis*, In Guide to big data applications (pp. 49–82), Springer International Publishing AG.

Damiani A., Vallati M., Gatta R., Dinapoli N., et al. (2015), *Distributed learning to protect privacy in multi-centric clinical studies*, In Conference on artificial intelligence in medicine in europe 65– 75.

Deist T.,Jochems A.,van Soest J., Nalbantov G.,et al.(2017) *Infrastructure and distributed learning methodology for privacy-preserving multi-centric rapid learning health care: euroCAT*, 4, 24-31.

Gkoulalas-Divanis A., Loukides G., & Sun J.(2014). *Publishing data from electronic health records while preserving privacy: A survey of algorithms*, Journal of biomedical informatics, 50, 4–19.

Jochems A., Deist T.M., Van Soest J., Eble M., et al.(2016), *Distributed learning: Developing a predictive model based on data from multiple hospitals without data leaving the hospital–a real life proof of concept*, Radiotherapy and Oncology, 121(3), 459–467.

Korn D.(2002), *The effect of the new federal medical-privacy rule on research*, The New England journal of medicine, 346(3), 201.

Lambin P., Zindler J., Vanneste B.G.L., Van De Voorde L., et al.(2016), *Decision support systems for personalized and participative radiation oncology,* Advanced Drug Delivery Reviews, 109, 131-153.

Lambin P.,Roelofs E.,Reymena B.,Velazquez E.R.,Buijsen J., et al. (2013), 'Rapid Learning health care in oncology' – An approach towards decision support systems enabling customised radiotherapy', 109, 159-164.

Lu C.L., Wang S., Ji Z., Wu Y., Xiong L., et al.(2015), *Webdisco: a web service for distributed cox model learning without patient-level data sharing*, Journal of the American Medical Informatics Association, 22(6), 1212–1219.

Meldolesi E.,van Soest J.,Alitto A.R.,Autorino R., et al.(2014), *VATE: VAlidation of high TEchnology based on large database analysis by learning machine*, 3(5), 435-450.

Patel V.L. & Kannampallil T.G.(2014), *Cognitive informatics in biomedicine and healthcare,* 53, 3-14.

Steyerberg E.W.(2008). *Clinical prediction models: a practical approach to development, validation, and updating*. Springer Science & Business Media.

Tagliaferri L., Kovács G., Autorino R., Budrukkar A., et al.(2016), *ENT COBRA(consortium for brachytherapy data analy- sis): interdisciplinary standardized data collection system for head and neck patients treated with interventional radiotherapy(brachytherapy)*, Journal of contemporary brachytherapy, 8(4), 336.

Je-LKS
Journal of e-Learning and Knowledge Society
The Italian e-Learning Association Journal
PEER REVIEWED PAPERS
Vol. 14, n.1, 2018
ISSN: 1826-6223 | eISSN: 1971-8829

EVALUATING INNOVATION INJECTION INTO EDUCATIONAL CONTEXTS

Francesca Pozzi[1]
Donatella Persico[1]
Luigi Sarti[1,2]

[1] Consiglio Nazionale delle Ricerche, Istituto per le Tecnologie Didattiche, Genoa - Italy
[2]SMART Engineering Solutions & Technologies (SMARTEST) Research Centre, eCampus University, Novedrate CO - Italy
pozzi@itd.cnr.it

Keywords: Technology acceptance; Training programme evaluation; Technology Enhanced Learning (TEL); Technology Acceptance Model (TAM); Teacher Professional Development (TPD).

One of the big challenges faced by research in the Technology Enhanced Learning (TEL) field has to do with the injection of innovation into real educational contexts. Very often, innovative technologies fail to be taken up by practitioners because of difficulties in absorbing both methodological and technological innovation of the target contexts. This may be caused by resistance of the target users associated with conservatism of the contexts, but also by inadequate approaches to innovation promotion or even lack of evidence of the return of investment of the innovation itself. Thus, a crucial need of the TEL field consists in the ability to evaluate both the efficacy of a new technology in the specific context to permeate, and the effectiveness and adequacy of the intervention designed to inject this innovation into the intended situation. This paper contributes to fill in this gap by proposing an approach that joins aspects of Guskey's model to evaluate the effectiveness of teacher training events together with

for citations:
Pozzi F., Persico D., Sarti L. (2018), *Evaluating Innovation Injection into Educational Contexts,* Journal of e-Learning and Knowledge Society, v.14, n.1, 83-95. ISSN: 1826-6223, e-ISSN:1971-8829
DOI: 10.20368/1971-8829/1302

indicators of the well-known Technology Acceptance Model, generally used to predict acceptance of a new technology. The approach proposed, called T&EAM (Technology & Event Acceptance Model), is illustrated. The discussion concerns its strengths and weaknesses and provides inputs for future applications and research.

1 Introduction

In the Technology Enhanced Learning research field, many projects aim to develop and inject methodological and technological innovation into a 'virgin' educational context. This process of exogenous (i.e. externally-driven) educational innovation usually leverages on teachers and is typically triggered by training events aiming to familiarize them with the technology, and consolidated by some kind of follow up, where they are scaffolded and guided through their first steps in the use of the new technology in real life contexts. In these situations, policy makers and/or researchers need to evaluate the results of both actions, thus assessing the adequacy and effectiveness of the training event as well as its effects in terms of longer term adoption of technology.

In this paper, we propose a 'joint approach', called T&EAM (Technology & Event Acceptance Model), built upon the conjunction of two existing and consolidated models, which have been merged to form a single framework for the evaluation of a training event and the associated technology-based educational innovation. The need for such an integrated approach derives from the awareness that the evaluation of a project's outcomes cannot be limited to the mere sum of the evaluation of the training event and of its practice-based follow-up. Rather, when the two actions are carried out in a synergic way, their evaluation too must be able to capture their joint effects, in order to fully appreciate the project outcomes.

We therefore intend to set the basis for the development of a framework that can be adopted in many other TEL projects, provided that they share the need of evaluating the effects of an innovation being injected into a new context in conjunction with a training initiative.

2 Theoretical background

The issue we intend to address here is the definition of an approach to evaluate the combined effects of the introduction of a new technology in a given context (and its methodological underpinnings) and of a training event addressing the perspective users. Our literature review therefore focuses on both aspects of the problem: the evaluation of the impact of a new technology in a given context and the evaluation of training events/programmes, and specifically those that aim to improve a teaching and learning process.

Both of these areas are very rich: there is plenty of models and frameworks addressing these issues, some of which are very well-known and consolidated. In the following sections we firstly concentrate on some of the most popular models to evaluate the impact of technological innovation, and secondly on the evaluation of training programmes.

2.1 Models for technology impact evaluation

A number of models have been proposed in the last decades to analyse and predict user acceptance of new technological tools (Davis, 1989; Rogers, 2010; Thompson *et al.*, 1991; Venkatesh & Davis, 2000).

Among these, some of the most well-known aim to predict users' intentions towards technology, and actual usage of it, as dependent variables, on the basis of various determinants (i.e. independent variables) that include: attitudes, perception of usefulness, perception of ease of use, motivation (both extrinsic and intrinsic), and other social factors. One of the most popular, the Technology Acceptance Model (TAM) (Chuttur, 2009; Davis, 1989), focuses on two determinants, *Perceived Usefulness* and *Perceived Ease of Use*, and has given rise to several derivatives and evolutions, often used in educational contexts (Cheung & Vogel, 2013; Liu *et al.*, 2010; Persico *et al.*, 2014; Tarhini *et al.*, 2013). For example, TAM2 (Venkatesh & Davis, 2000), considers some additional determinants concerning social influence, including for example *Subjective Norm*, defined as "the person's perception that most people who are important to him think he should or should not perform the behaviour in question" (Fishbein & Ajzen, 1975, p. 302). As described in the following, TAM and TAM2 provide the foundations for the development of our evaluation approach, although the three variables (*Perceived Usefulness*, *Perceived Ease of Use* and *Subjective Norm*) are not used as determinants, to predict behaviour, but as indicators of acceptance, after usage of the technology.

2.2 Models for training initiatives evaluation

As mentioned above, there is a multiplicity of models and studies dedicated to the evaluation of training programmes and training initiatives of different kind. With no intention to be exhaustive, we examine here those that have inspired our approach.

The Kirkpatrick's 4 levels model is probably one of the most well-known and widely applied. It considers 4 levels of training evaluation: reaction (a measure of satisfaction of the people involved in the training initiative), learning (a measure of knowledge and skills increase), behaviour (a measure of change in behaviour) and results (a measure of the effects on the institutions) (Kirkpatrick, 1994).

Guskey's 5-level model is also an extension of Kirkpatrick's, with the peculiarity of having been adapted to a teacher training context, thus paying special attention to effects on school contexts and students. It encompasses the following levels: participant reaction, participant learning, organizational support and learning, participant use of new knowledge and skills, student learning outcomes (Guskey, 2000).

3 The T&EAM approach

This section describes the T&EAM approach, the associated indicators, as well as the tools to be used for data collection.

3.1 Evaluating the technology acceptance with the T&EAM approach

As already mentioned, the TAM and its subsequent evolutions were chosen as the backbone approach to evaluate the technology in the T&EAM approach, even if it is acknowledged that this model was originally devised as a predictive tool. However, Persico *et al.* (2014) have already shown how the TAM indicators "perceived ease of use" and "perceived usefulness" can be used for ex-post assessment of the impact of a technology, by collecting information concerning users' opinions about these two indicators and complementing them with data gathered from other sources, such as observation and data tracked by the system itself. Furthermore, the subjective norm indicator introduced by TAM2 is also used.

The reasons for the choice of TAM and TAM2 indicators (Venkatesh & Davis, 2000) as main indicators of the T&EAM approach are two-fold: first, the number of experiences and studies where they have been applied witness their capacity to adapt to several contexts, including teachers' acceptance of technology (Huntington & Worrell, 2013; Persico *et al.*, 2014). Especially in those studies concerning the barriers to technology uptake by teachers (Delfino *et al.*, 2004; Lambert *et al.*, 2008; Lloyd & Albion, 2009), the TAM indicators have proved to be key determinants. Thus, training initiatives that can improve some of these factors are more likely to increase the chances that the proposed technology is adopted in the long run.

A second reason for this choice is that these models are applicable to any technology, if their indicators and the evaluation means are tailored to the system structure, functions and user types. This process of adaptation/tailoring is essential, especially when dealing with formative evaluation, as the accuracy of the problems' diagnosis improves with it.

Thus, in our approach the "perceived ease of use" and "perceived usefulness" indicators are used to build data collection tools aiming to understand the users' opinions after use of the technology during ad hoc training event(s). In our model, these subjective data are complemented with more objective data about actual usage of the system. This latter information is typically obtained thanks to tracking mechanisms built in the technology, usually with learning analytics techniques (Persico *et al.*, 2014). These data provide, among other things, a measure of trustworthiness of the users' opinions. If, for example, a user says that a given functionality was easy to use, but tracked data show he/she never used it, his/her opinion is less trustworthy than that of a user who claims the functionality was difficult to use after having engaged with it for a significant amount of time.

3.2 Evaluating the workshops with the T&EAM approach

In the proposed approach, the evaluation of the training initiative(s) used to introduce the technology in one context is carried out according to Guskey's model (2002). This model is derived from Kirkpatrick's work (1994); evidence is collected and analysed at five critical levels:

1. workshop participants' reactions (i.e. perceptions on the training event);
2. workshop participants' learning (i.e. knowledge and skills gained);
3. organization support and change (i.e. impact on the organization where the participants work and organisation's support to the implementation of the innovation);
4. participants' use of new knowledge and skills (i.e. application of the acquired competence in the teaching profession);
5. student learning outcomes (i.e. impact on the students who are the ultimate beneficiaries of the innovation proposed).

While most evaluation models focus on levels 1 and 2, Guskey's model also takes into consideration factors that can facilitate or hinder innovation within an organization (level 3) and long term effects of the training events on participants (level 4), as well as on their students (level 5). This is the main benefit of this model in respect to the others.

According to the T&EAM, while level 1 to 3 are typically gauged at the end of the training event(s), level 4 and 5 data collection takes place after the follow up (medium term). The data collected from training participants are also complemented with data concerning the actual training sessions. These data are typically collected during the events by an observer, taking notes on the basis of a rubric.

3.3 Data collection process

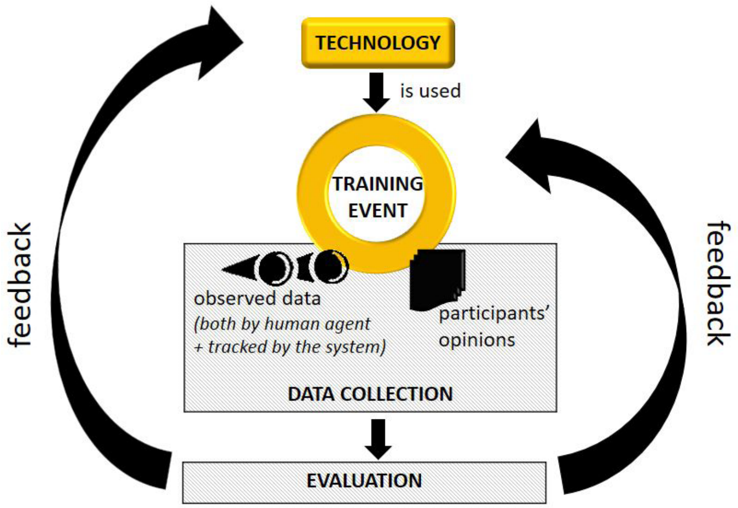

Fig. 1 - The T&EAM evaluation approach

Overall, in the T&EAM approach we have merged the TAM and the Guskey's models, customized their original indicators, and created joint evaluation means, thus forming a unique evaluation framework for data collection and data analysis.

The resulting T&EAM approach (see Fig. 1) allows to strike a balance between the need to carry out a deep analysis and evaluation of different aspects of the technology and the training events, on one hand, and the requirement to keep the effort of the users relatively low, so to make the approach more sustainable.

Fig. 1 represents the cyclic process of data collection and evaluation providing feedback on both the technology and the teacher training events. The data collected concern:

- Participants' opinions, gathered at the end of the training event(s), in a very easy and relatively unobtrusive way through questionnaires and interviews;
- Actual participants' behaviour during the events, annotated by human observers and/or automatic tracking.

The complete list of indicators is reported in Table S1 of the Supplementary file 1.

3.4 Managing evaluation within projects

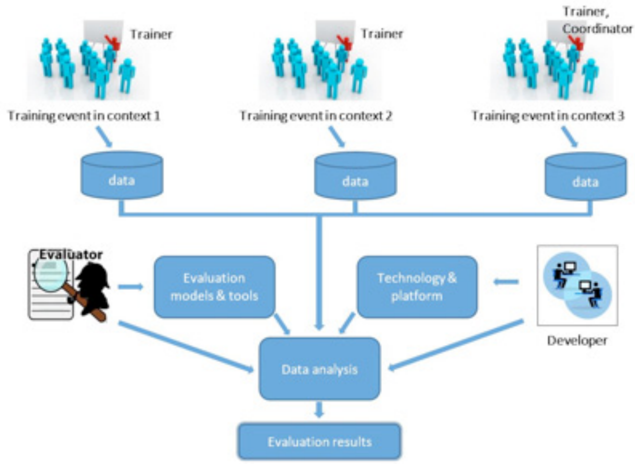

Fig. 2 - Collaborative evaluation of training events according to the T&EAM approach

Boosting innovations into real contexts, in the context of complex (European) projects, is often done through several parallel events held in different locations and require data to be collected in a homogenous and comprehensive way (Pozzi *et al.*, 2015). The actors usually involved in projects of this kind, typically comprise (see Fig. 2):

- a number of institutions/agents that carry out the pilot of the training events in one or more contexts (indicated as the trainers, in the following);
- one institution leading the evaluation (identified as the evaluator in the following);
- one actor in charge of the development and tuning of the technology (the developer).

The coordinating institution (the coordinator) could be any of the above actors, except the evaluator, to avoid conflicts of interest. The evaluator usually devises or instantiates the evaluation model, designs and produces the evaluation tools, coordinates data collection (which is carried out on site by the trainers) and carries out the data analysis (see Fig. 2).

In case the evaluation involves institutions in different countries, language problems need to be handled with the support of local partners; so, for example, the questionnaires should be developed in one common language (typically

English), and translated into the local languages. A first phase of analysis of any narrative (answers to open questions or interviews) should be carried out by the trainers, based on common guidelines provided by the evaluators, to produce raw data in English that can easily be interpreted by the evaluators.

4 Discussion

The T&EAM approach has been developed and experimented for the first time in METIS[1], a LLP Project aiming to foster methodological and technological innovations in learning design. In this project, the authors of this paper where in charge of the evaluation workpackage (Pozzi *et al.*, 2013; 2015a; 2015b). Within METIS, the target of the innovation were three different educational contexts (namely Higher Education, Vocational Training and Adult Education), thus the evaluation approach was applied to these three situations. Indeed, the T&EAM approach proved flexible enough to fit in with the three different contexts, and appears to be potentially exportable to several other educational contexts (Pozzi *et al.*, 2015).

The evaluation means were questionnaires and interviews based on rubrics produced in English by the evaluator and translated in Spanish and Greek by the local partners. A first round of the qualitative analysis was carried out locally, to produce English narratives corresponding to the open answers to questionnaires and interview transcripts.

Within the METIS project the application of the T&EAM evaluation approach yielded a wealth of information about ease-of-use, usefulness and actual use of the innovative technological system proposed to teachers (Asensio-Pérez *et al.*, 2017). These information referred specifically to the various functionalities of the technological system introduced by METIS, so the project partners were able to use them to improve and tune both the proposed technology and the training format, thus increasing the possibility that the technology is taken up by other actors in the same (or similar) contexts.

The approach allowed us to collect the data in a very unobtrusive way, with data collection carried out by the project partners in charge of the training in each context according to the guidelines provided by the evaluators.

This organization allowed for the T&EAM approach to be easily and consistently adopted and managed even by the partners who were not directly involved in its conceptualization. In particular, the online questionnaire proved to be very easy to be managed, once translated in the local languages; the interviews, carried out by the local partners and based on a common rubric provided in English, were slightly more complicated, because they required a certain amount of time and an effort to produce a synthesis in English of the

1 http://www.metis-project.org/

interviewee answers. Data collection through interviews was possible as long as the number of interviewees is relatively small; in case of big numbers, probably they should be replaced by questionnaires or even group data collection techniques, such as focus groups.

As far as the indicators are concerned, the ones deriving from the TAM model and devoted to evaluating technology acceptance provided information about ease of use and usefulness of each individual function of the software. Given that in METIS the number of functions implemented in the technology was very high, in order to make it easier for respondents to recall the functions referred to by the individual questions in the questionnaire, these were enriched with pictures of the platform, so to highlight the interface controls associated to the various program functions. This proved to be an effective strategy that allowed the users to straightforwardly understand the questions.

The indicators focusing on the training coming from Guskey's model were also very useful: not only did they yield information about the adequacy of the workshops in the different contexts, but they also informed us about the possibility that the technology would really be taken up in the various situations. Some problems emerged when collecting data about the Student Outcomes indicators, as it often happens in TEL research, because evidence about students learning appears very difficult to assess, as innovative methods and technologies cannot be easily compared with traditional ones. Probably, structured data collection protocols would have helped teachers to systematically collect more significant data about students learning ad this is something that should be fixed for future adoption.

Another challenge posed by the T&EAM approach regarded the juxtaposition of the data tracked by the system and those coming from the questionnaires and interviews. One of the reasons for these difficulties is the difference of granularity between the data typically tracked by the platform and those collected through the questionnaires and interviews. While the former are usually low level data, concerning individual actions of the users, the latter are higher-level data referring to the technology functions. Their comparison might require some effort to elaborate and aggregate the tracked data, so that they can be used to put in the right light the users' opinions on the technology functions.

As a last consideration, we should note that, usually the life span of a project is rather short and does not allow to wait for long term evidence that the innovation really permeates the target system. As a consequence, what can realistically be evaluated is the acceptance of the technology, the impact of the training event, as well as the short/medium term changes compared to the original conditions of the target context(s).

Conclusions

The T&EAM evaluation approach presented above aims to assess the acceptance of an innovative technology, when this is introduced for the first time into an educational context through some kind of training programme.

The novelty of the model lies not so much in the indicators and tools used, which mainly derive from other existing and well-known evaluation models, but rather in the way they are used and integrated into one coherent evaluation framework thus producing an overarching model. The proposed evaluation means jointly assess the technology and the training events and consider all the variables that may affect the uptake of the innovation, in order to produce a comprehensive picture of the forces that may foster or hinder the integration of the innovation into real conditions.

Even if the T&EAM has been conceived in the framework of one specific project, we believe the problem addressed is frequent in the TEL field, where many of the projects funded by the EC or other funding agencies aim to introduce methodological and technological innovation into established educational systems. For this reason, further research directions should aim to investigate transferability to projects with similar intents.

As to the authors, further research efforts will be devoted to the identification of the invariant factors of the model and of the degrees of freedom left to the evaluators when applying the model.

Acknowledgments

This study was funded by Project METIS (*Meeting teachers co-design needs by means of Integrated Learning Environments*) – EC Lifelong Learning Programme (Project Number: 531262-LLP-1-2012-1-ES-KA3-KA3MP. Agreement n° 2012-3971/001-001).

The authors declare that they have no conflict of interest.

Data are available upon request to the corresponding author.

REFERENCES

Asensio-Pérez, J. I., Dimitriadis, Y., Pozzi, F., Hernández-Leo, D. Prieto, L. P., Persico, D., & Villagra-Sobrino, S. L. (2017), *Towards teaching as design: Exploring the interplay between full-lifecycle learning design tooling and Teacher Professional Development*. Computers & Education, 114, 92-116.

Cheung, R., & Vogel, D. (2013), *Predicting user acceptance of collaborative technologies: An extension of the technology acceptance model for e-learning*. Computers & Education, 63, 160–175. doi:10.1016/j.compedu.2012.12.003

Chuttur, M. (2009), *Overview of the Technology Acceptance Model: Origins, Developments and Future Directions*. Sprouts: Working Papers on Information Systems, 9(37).

Davis, F. D. (1989), *Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology*. MIS Quarterly, 13(3), 319–339. doi:10.2307/249008

Delfino, M., Manca, S., Persico, D., & Sarti, L. (2004), *Online learning: attitudes, expectations and prejudices of adult novices*. In V. Uskov (Ed.), Proceedings of the IASTED Web-Based Education Conference (pp. 31–36). Calgary, Canada: ACTA Press.

Fishbein, M., & Ajzen, I. (1975), *Belief, attitude, intention and behavior: an introduction to theory and research*. Reading, Ma: Addison-Wesley Pub.Co.

Guskey, T. R. (2000), *Evaluating Professional Development. Thousand Oaks*: Corwin Press.

Guskey, T. R. (2002), *Professional Development and Teacher Change*. Teachers and Teaching, 8(3/4), 381–391. doi:10.1080/135406002100000512

Huntington, H., & Worrell, T. (2013), *Information Communication Technologies in the Classroom: Expanding TAM to Examine Instructor Acceptance and Use*. Journal of Educational Multimedia and Hypermedia, 22, 147–164.

Kirkpatrick, D. L. (1994), *Evaluating training programs: The four levels*. San Francisco, CA: Berrett-Koehler.

Lambert, J., Gong, Y., & Cuper, P. (2008), *Technology, Transfer and Teaching: The Impact of a Single Technology Course on Preservice Teachers' Computer Attitudes and Ability*. Journal of Technology and Teacher Education, 16(4), 385–410.

Liu, I.-F., Chen, M. C., Sun, Y. S., Wible, D., & Kuo, C.-H. (2010), *Extending the TAM model to explore the factors that affect Intention to Use an Online Learning Community*. Computers & Education, 54(2), 600–610. doi:10.1016/j.compedu.2009.09.009

Lloyd, M., & Albion, P. (2009), *Altered Geometry: A New Angle on Teacher Technophobia*. Journal of Technology and Teacher Education, 17(1), 65–84.

Persico, D., Manca, S., & Pozzi, F. (2014), *Adapting the technology acceptance model to evaluate the innovative potential of e-learning systems*. Computers in Human Behavior, 30, 614–622. doi:10.1016/j.chb.2013.07.045

Persico, D., Manca, S., & Pozzi, F. (2014), *Adapting the technology acceptance model to evaluate the innovative potential of e-learning systems*. Computers in Human Behavior, 30, 614–622. doi:10.1016/j.chb.2013.07.045.

Pozzi, F., Ceregini, A., Persico, D., & Sarti, L. (2015a), *METIS D5.3: Report on second formative evaluation round*. METIS project deliverable.

Pozzi, F., Ceregini, A., Persico, D., & Sarti, L. (2015b), *METIS D5.4: Final evaluation report*. METIS project deliverable

Pozzi, F., Persico, D., & Sarti, L. (2013), *D5.1 Assessment plan*. METIS project deliverable

Pozzi, F., Persico, D., & Sarti, L. (2015), *Evaluating The Acceptance Of An Innovative Learning Design Environment Within Communities Of Practitioners*. Procedia -

Social and Behavioral Sciences, 191, 1019 – 1023..

Rogers, E. M. (2010), *Diffusion of Innovations* (4rt ed.). New York, NY: Simon and Schuster.

Tarhini, A., Hone, K., & Liu, X. (2013), *User Acceptance Towards Web-based Learning Systems: Investigating the Role of Social, Organizational and Individual Factors in European Higher Education*. Procedia Computer Science, 17, 189–197. doi:10.1016/j.procs.2013.05.026

Thompson, R. L., Higgins, C. A., & Howell, J. M. (1991), *Personal Computing: Toward a Conceptual Model of Utilization*. MIS Quarterly, 15(1), 124–143. doi:10.2307/249443

Venkatesh, V., & Davis, F. D. (2000), *A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies*. Management Science, 46(2), 186–204. doi:10.1287/mnsc.46.2.186.11926

Je-LKS
Journal of e-Learning and Knowledge Society
The Italian e-Learning Association Journal
PEER REVIEWED PAPERS
Vol. 14, n.1, 2018
ISSN: 1826-6223 | eISSN: 1971-8829

CONVERSATIONAL FUNCTIONS FOR KNOWLEDGE BUILDING: A STUDY OF AN ONLINE COURSE AT UNIVERSITY

Stefano Cacciamani[1]
Vittore Perrucci[1]
Ahmad Khanlari[2]

[1]University of Valle d'Aosta
[2]University of Toronto
s.cacciamani@univda.it; v.perrucci@univda.it; a.khanlari@mail.utoronto.ca

Keywords: Knowledge Building, online environments, collaborative learning, Conversational Functions.

Interactions in online courses have been studied by analyzing Conversational Functions used by participants. Cacciamani, Perrucci Khanlari (2016) developed a coding scheme, named CF4KB, consisting of four Global Conversational Functions (GCF), each articulated in two Specific Conversational Functions (SCF). The aims of the present study were to explore: 1) What are the more frequently SCF used by the participants, both at the beginning and at the end of an online course, and if there are differences between the beginning and the end of the course in terms of SCF used; 2) If there is any specific pattern of SCF used at the beginning and at the end of the course and if there are any persistences in using the same SCF. For these aims, 152 messages posted in Knowledge Forum online environment by 24 university students (19 F and 5 M) were considered. The messages have been segmented into units of meaning and the 1451 resulting segments have been coded by two independent judges who applied

for citations:
Cacciamani S., Perrucci V., Khanlari A. (2018), *Conversational Functions for Knowledge Building: a Study of an Online Course at University,* Journal of e-Learning and Knowledge Society, v.14, n.1, 95-107. ISSN: 1826-6223, e-ISSN:1971-8829
DOI: 10.20368/1971-8829/1391

the CF4KB. The analysis of frequencies evidenced the more frequently used SCF, at the beginning and the end of the course. Comparing SCF frequencies between the beginning and the end of the course, differences were detected. Results showed also different patterns in the use of the SCF at the beginning and at the end of the course. In addition, the persistence of one SCF was found. Implication of these results for the analysis of the interactions in online courses are discussed.

1 Introduction

The aim of Computer Supported Collaborative Learning (CSCL) is to embrace technology to facilitate collective activity and collaborative learning. The general belief about CSCL is that sharing knowledge between contributors as well as constructing individual's own knowing are primarily achieved through situated discourse processes (Stahl, 2003). Therefore, it is required for CSCL researchers to investigate the discourse processes, in order to understand substantial, complex, and interactive process of joint activities between the participants (Lipponen, Hakkarainen, & Paavola, 2004). These beliefs have encouraged CSCL community to focus on studying and analyzing discourses in CSCL environments. Of particular interest in this field are studies on Conversational Functions (CF) in collaborative learning environment. As described by Wise, Saghafian and Padmanabhan (2012), Hare (1994) has defined functions as particular rights or duties assigned to roles which guide role takers to interact with other community members and contribute to the conversation. Each role may have one or more functions that need to be fulfilled by the role taker in order to support a productive conversation. Building on this definition, Wise and colleagues defined CF as a

> "specific kind of activity performed in a discussion that is expected to support productive interaction"(Wise *et al*., 2012, p. 57).

In their study, the authors analyzed students' online interactions to understand what kinds of CF are performed by the participants during the joint activities and identified the following seven common CF performed by students: motivate others to contribute, give direction to the conversation, provide new ideas, use theory to ground the discussion, bring in (relevant external) sources, respond to previous comments, and summarize existing contributions**.**

Building on this work, Cacciamani, Perrucci and Khanlari (2016) have recently developed a coding scheme, called "Conversational Functions for Knowledge Building" (CF4KB). Knowledge Building (KB) is defined by Scardamalia and Bereiter (2003a) as the production and continual improvement of ideas of value to a community through social interactions. KB holds an even stronger belief in the role of discourse in learning:

> "the state of public knowledge in a community only exists in the discourse of that community, and the progress of knowledge just is the progress of knowledge-building discourse" (Scardamalia & Bereiter, 2006, p. 12).

As Bereiter and colleagues (1997) assert, students who fail to master knowledge-building discourse, have failed to master science.

Scardamalia and Bereiter (2006) described 12 principles for KB, including Real Ideas and Authentic Problems, Improvable Ideas, Rise Above, Epistemic Agency, Community Knowledge, Collective Responsibility, Constructive Uses of Authoritative Sources, Knowledge Building Discourse, and Concurrent, Embedded, and Transformative Assessment. These 12 principles frame KB as an idea-centered pedagogy with students as epistemic agents, creating knowledge through engaging in complex socio-cognitive interactions. Although CSCL environments and KB Environments are usually considered as synonym, Scardamalia and Bereiter (2003b) have articulated several features of KB environments which distinguish these two types of collaborative environments:

1. KB environments are self-organized, self-directed environments and support for advanced knowledge processes that contributors need;
2. in KB environments, collective knowledge advances built from the contributions of community members are represented in shared, user-configured design spaces;
3. in KB environments, contributions to the evolution of ideas are evident, as students cite and reference one another's work;
4. in KB environments, students represent higher-order organizations of ideas and show the rising status for improved ideas;
5. in KB environments, students consider different ways for the same idea to be worked with in varied and multiple contexts and to appear in different higher-order organizations of knowledge;
6. in KB environments, students provide feedback to enhance self-and group-monitoring of ongoing processes and to tap idea potential.

These are the essence of the 12 KB principles which describe Knowledge Building as a pedagogy which has the potential to "increase the likelihood that what the community accomplishes will be greater than the sum of individual contributions and part of broader cultural efforts" (Scardamalia & Bereiter, 2003a).

The implementation of the KB in an online course at University could help students to move from a model of work with knowledge centered on "acquisition of knowledge", towards a model centered on the "creation of knowledge" (Paavola & Hakkarainen, 2005).The first model assumes that knowledge is a property of an individual mind. This approach is easily connected to a 'folk theory' of mind according to which the mind of learners is a container of knowledge, and learning is a process that fills the container, implanting knowledge there by the teacher (Bereiter, 2002). The second model, assumes that knowledge is actively constructed by the learner collaborating whithin a community, and implies an active position for the students in the work with knowledge. Scardamalia and Bereiter (2006) describe this position in terms of Epistemic Agency.

The work of Cacciamani and colleagues (2016) focused on cognitive aspects of the knowledge building process, considering these characteristics of a genuine KB environment, in order to detect CF in online activities in Knowledge Forum (KF)®. KF is the most widely used KB environment to support collaborative knowledge creation. CF4KB scheme identifies four Global Conversational Functions (GCF) articulated in eight Specific Conversational Functions (SCF):

- The GCF of *Exploring* which is mapped to the "Real Ideas, Authentic Problems" and "Epistemic Agency" principles of KB model includes the following two SCF: A) *Question or problem of inquiry*, by which students propose questions or problem of inquiry concerning the course content, *B) Hypothesis and ideas,* by which students formulate possible explanations about a question or problem of inquiry emerging during online discussion.
- The GCF of *Providing information* which is mapped to the "Constructive use of Authoritative Recourses" KB principle includes the following two SCF*: C) Applicative examples*, by which participants provide examples according to their personal experience, *D) Information from authoritative sources*, by which the participants provide theoretical information that is explicitly referred to a source.
- The GCF *of Re-elaborating,* which is mapped to the Rise Above KB principle, includes the following SCF: *E) Repetition/Quotation others' idea*: formulating explicit reference to an idea of another member of the community, and *F) Synthesis*: formulating a synthesis using ideas of different participants.
- The GCF *of Evaluation* is mapped to the "Concurrent, Embedded and Transformative Assessment" and includes the following SCF: *G) Comment*: providing content evaluation, including judgments of agreement or disagreement, positive or negative on a content expressed by another participant, *H) Metacognitive reflection and Metacommunication*: providing evaluation or reflections on the strategies of work of the online course or metacommunication about the activity.

The current study employs this coding scheme to explore:

- What are the more frequently SCF performed by the participants in the activity of knowledge building, both at the beginning and at the end of an online course, and if there are differences between the beginning and the end of the course in terms of SCF used.
- If there is any specific pattern for using SCF at the beginning and at the end of the online course and if there are any persistence in using the same SFC from the beginning to the end of the online course.

2 Method

2.1 Participants and course background

The dataset used for this study is comprised of students' discourse in the online course of Educational Psychology, as archived in Knowledge Forum® (KF), an environment built specifically to support production and refinement of the community's knowledge. The course was held in 2006-2007 academic year at the University of Valle d'Aosta and was organized in four modules. The participants included 26 (21 F and 5 M) undergraduate students of Primary Education, as well as undergraduate students of Educational Sciences, in addition to the teacher and tutor.

For research purposes, 152 messages posted by 24 students (19 F and 5 M), who wrote in both Module 1 (= 83 notes) and Module 4 (= 69 notes) were considered[1].

2.2 Procedure

The procedure, inspired by Strijbos and colleagues (2006), included the following steps: 1. develop the segmentation procedure according to the unit of analysis; 2. apply the coding scheme categories; 3. determine the agreement of the codification; 4. resolve the controversial cases. The 152 messages have been segmented into units of meaning, identified by punctuation (i.e. full stops, suspension dots, exclamations, and question marks), used by the author of the message, (cf. Strijbos *et al.*, 2006). The 1451 resulting segments have been coded by two independent judges who applied the CF4KB coding scheme at SCF level. The overall inter-agreement amounted to 77.3%, with a K of Cohen = 0.66, considered good in the literature (Landis & Koch, 1977). Controversial cases were discussed until the complete agreement by the same two judges.

[1] Two students that had not posted neither in Module 1 nor in Module 4 were not considered.

2.3 Data Analysis

In order to answer the first research question, the percentage of each SCF was separately computed for both Module 1 and Module 4. Then, employing Wilcoxon Test, Module 1 and Module 4 were compared, in terms of the frequencies of each SCF.

In order to explore the second research question, for each SCF, correlations between the frequencies were conducted, by means of Spearman's statistic, both within each Module and between Module 1 and Module 4.

3 Results

3.1 Research Question 1

In Table 1, for each SCF used in Module 1 and Module 4, frequencies and percentages of the total of segments are reported.

Table 1
FREQUENCIES (F) AND PERCENTAGES (%) OF THE PERFORMED SCF IN MODULE 1 AND MODULE 4

SCF	Module 1 *f (%)*	Module 4 *f (%)*
A. Question or problem of inquiry	7 (0.78)	12 (2.19)
B. Hypothesis and ideas	246 (27.27)	243 (44.26)
C. Applicative Examples	13 (1.44)	66(12.02)
D. Information from authoritative sources	547 (60.64)	165 (30.05)
E. Repetition/Quotation others' idea	6 (0.67)	0 (0)
F. Synthesis	4 (0.44)	1 (0.18)
G. Comment	12 (1.33)	9 (1.64)
H. Metacognitive reflection and Metacommunication	12 (1.33)	19 (3.46)
I. Other	55 (6.10)	34 (6.19)
TOTAL	902 (100)	549 (100)

As Table 1 shows, the more frequently used SCF in both Module 1 and Module 4 are *Hypothesis and ideas* (B) and *Information from authoritative sources* (D), while the less frequently used SCF are *Synthesis* (F) and *Repetition/Quotation others' idea* (E), which is completely absent in the last module. Nevertheless, from Table 1 it is evident that the *Information from authoritative sources* (D) is more frequent in Module 1 than in Module 4. From Wilcoxon Test this difference results significative ($Z = -2.68$, $p <.01$). Moreover, Wilcoxon test results evidence a significant difference ($Z = -3.22$, $p <.01$) between frequencies of *Applicative examples* (C); in Module 4 students used more *Applicative examples* compared with Module 1 (see Table 1).

3.2 Research Question 2

In Table 2, correlation between the SCF are separately shown for both Module 1 and Module 4.

Table 2

CORRELATIONS (RHO) BETWEEN THE SCF IN MODULE 1 AND MODULE 4

SCF	A	B	C	D	E	F	G	H	I
A		.041	-.254	-.141	-.257	-.093	-.308	-.050	-.286
B	-.037		.564**	-.184	-.307	.091	.269	.320	.431*
C	-.208	.048		.174	.082	.278	.638**	.378	.586**
D	-.334	-.061	.041		-.028	.076	.149	.157	.093
E	-	-	-	-		-.120	.333	-.082	.043
F	.456*	.349	-.247	.181	-		.217	.377	.203
G	-.050	.209	.027	-.071	-	.436*		.257	.449*
H	.144	.408*	.082	.048	-	.394	.199		.390
I	.050	.227	.034	.263	-	.359	.196	.787**	

Module 4 (below diagonal); Module 1 (above diagonal)

* $p < .05$; ** $p < .01$

As shown, in Module 1, the SCF of *Applicative examples* (C) significantly correlates with *Hypothesis and ideas* (B) (Rho = .564, $p < .01$), *Comment* (G) (Rho = .638, $p < .01$) and *Other* (I) (Rho = .586, $p < .01$). Although the correlation between *Hypothesis and ideas* (B) and *Comment* (G) is not significant, all the correlations between *Other* (I) and *Hypothesis and ideas* (B), *Applicative example* (C), *Comment* (G) are significant (Rho = .431, $p < .05$; Rho = .586, $p < .01$; Rho = .449, $p < .05$, respectively).

In Module 4, the SCF of *Synthesis* (F), significantly correlates with both *Question or problem of inquiry* (A) (Rho = .456, $p < .05$) and *Comment* (G) (Rho = .436, $p < .05$). Moreover, the correlation betweeen *Hypothesis and ideas* (B) and *Metacognitive reflection and Metacommunication* (H) appears as significant (Rho = .408, $p < .05$).

Frequencies correlations between Module 1 and Module 4 run for each SCF, show a statistical significance only for *Hypothesis and ideas* (B) (Rho = .58, $p < .01$).

Conclusion

The present study was focused on two questions of inquiry. The first one was oriented to identify the more frequently SCF performed in the activity of knowledge building by the participants, at the beginning and at the end of the online course. We explored also the differences in SCF performed in the first and last part of the course, if there is any. The second question of inquiry was oriented to analyze if there is any specific pattern of using SCF at the beginning and at the end of the online course. In addition, we explored if there are any persistences in using the same SFC from the beginning to the end of the online course.

Concerning the first question of inquiry, results showed that both in the first and in the last module of the online course, the more frequently SCF used are *Information from authoritative sources* and *Hypothesis and ideas,* respectively. Moreover, the less SCF used are *Synthesis* and *Repetition/Quotation others' idea.* It can indicate that the students use the online environment as a place to share their explorative activity (corresponding to the GCF of *Exploring*) focused on building hypotheses and ideas about the problems discussed and on providing information from authoritative sources. In this work, the use of the others' ideas of the and also the synthesis of ideas developed in the online discussion (corresponding to the GCF of *Re-elaboration*) are not so frequent. These results are consistent with a study conducted by Pena-Shaff and Nicholls (2004), showing that in the online interactions analyzed in a bulletin board system used by university students, few messages provided a summary of the ideas presented in a discussion thread; therefore, most discussion seemed to be left unfinished. This result can indicate that the students are not aware of the possibility to improve the common knowledge using other's ideas in combination with their own ideas and the sources of information. They also seem not to perceive the relevance to create a synthesis helping the common systematization of the knowledge advancement developed.

Comparing the first and the last module, we have seen also a decrease in the use of *Information from authoritative sources* concomitant with an increase of *Applicative examples*. This result can indicate the progressive abandonment of the working model, centered on "acquisition of knowledge", towards the model centered on the "creation of knowledge" (Paavola & Hakkarainen, 2005). The students seemed to assume the first model in the first part of the course, organizing their online activity essentially in terms of writing in their notes the information derived from the handbook. In the last part of the course, students seemed to adopt the second model, assuming an Epistemic Agency expressed by the reduced impact in the online activity of the information derived from the handbook (comparing the number of segments of this SCF in the first and last module). Students connect, indeed, actively theoretical concepts derived from handbook with examples from their own experience and used them to create or elaborate ideas.

With reference to the second question of inquiry results showed two patterns in the use of the SCF emerging at the beginning of the course. These patterns are indicated by the correlation between *Hypothesis and ideas* and *Applicative Example,* as well as correlation between *Applicative Example* and *Comment,* in Module 1. The first pattern can be considered from one hand as having an argumentative purpose: students who proposed hypotheses or ideas, also provided examples from their experiences to support them. From the other hand, it can be also interpreted as having a reflective purpose; students who described examples from their experience, can reflect on them and create hypotheses and ideas. Also, in the second pattern we can see argumentative or reflective purposes: some students produced comments on the note contents of other members and provided examples to support their own comments. Otherwise, students described in the online activity personal experiences and reflecting on them, identified ideas to make comments on the notes contents of the other community members. In the patterns of the first module, also the SCF of *Other* is present, through a statistical significant correlation with *Hypothesis and ideas*, *Applicative examples*, and *Comment*. This can indicate that students involved in these patterns, also performed SCF not necessarily focusing on knowledge building activity.

Three different patterns emerged in Module 4. The correlation between *Question or problem* of inquiry with *Synthesis*, seems to indicate from one hand that being more active in posing questions is associated with summarizing the ideas emerged in the online discussion. From the other hand, it can also indicate that the synthesis can be the starting point for new questions, showing that students are deepening their understanding. In addition, the correlation between *Hypothesis and ideas* and *Metacognitive reflection and Metacommunication* (corresponding to the two GCF of *Exploring* and *Evaluating*) can indicate that, at the same time, the community members are paying attention to the exploration and evaluation activity. From one hand, in fact, to seek explanations about the problems and questions of inquiry can lead to reflect on the strategy of work used in the knowledge building activity. From the other hand, this kind of reflection can stimulate students agency toward the creation of new hypotheses and ideas about the problems discussed. This seems to indicate that students are more aware that they are working towards the collective goal of building common knowledge. Finally, the correlation between *Synthesis* and *Comment* shows that the need to understand what the advancements of community knowledge are -expressed in making a synthesis- is associated with the interaction with the ideas contained in the messages of the other community members. This correlation shows that the expression of comments on the others' ideas can stimulate producing a synthesis, probably perceived as a "tool" to trace the common effort to build knowledge, expressed in reciprocal comments. It is interesting to note that, unlike Module 1, there is no correlation between the SCF of these patterns and the category *Other*: students involved in the patterns of Module 4 tend to not to use SCF not implied in knowledge building activity.

These three patterns can help to understand how the online interaction between students works and to describe in which ways they are assuming Epistemic Agency about the work with knowledge in the online course (Scardamalia & Bereiter, 2006).

With reference to the correlation between Module 1 and Module 4, we have found that introducing hypotheses and ideas in Module 1 is associated with the use of the same SCF in Module 4. This can indicate that there are some students who are more capable or available to work with hypotheses and ideas, compared to other students. Combining the present result with the previous ones, these students in the first part of the course connect the activity of *building hypotheses and ideas* with *providing examples*, as well as with SCF of *Others*, which is not oriented toward knowledge building activity. In the last part of the course, *building hypotheses and ideas* is associated with the *metacognitive reflection/communication*. We can interpret this result in terms of an "emerging role" (Strijbos & Weinberger, 2010). This construct highlights that the group work members develop spontaneously during their collaborative learning activity roles that helps the group to work effectively. In this respect, the presence of students who tended to introduce hypothesis and ideas in Module 1 and Module 4 can be considered in terms of an "emerging role", helping other students to work on creating new ideas and improving them. This role seems to be associated with two different patterns: it moves from a personal argumentative or reflective purpose in the first module towards a more collective purpose focused on combining cognitive elaboration and strategies to be used in the community for knowledge building activities. This role can favor the students' transition from the "acquisition model" to the "construction model" in the work with knowledge.

In terms of limitations, all the participants were from a specific context (the same university) and it may limit the ability to generalize the results. Therefore, the development of this research needs to overcome this limitation, by other experimental tests to give statistically more support to the results and estimate the extent to which they are generalizable (with different context, skill level of participants, topic, type of conversation, etc.).

Despite this limitation, the present study can offer a relevant contribution in terms of both knowledge advancements for the research in the field and new direction of inquiry.

In terms of new contribution for knowledge advancement, the results showed, using the SCF analysis, that students involved in an online course where KB has been implemented, moved really from an "acquisition model" toward a "construction model" in the work with knowledge. The analysis through the CF4KB coding scheme allowed to identify this change of model analysing, in particular, how students performed the SFC, how their use changed during the online course, which kind of patterns in the SCF use and which role emerged in the online activity.

New directions of the inquiry could be, then, focusing, in the design of online courses at University, on the study of the conditions that can favor the change of the students' model of work with knowledge, using the CF4KB coding scheme to detect this change. First of all it is possible to study the relationships between the change in the use of SCF, during the online activity, with reference to the aspects of the course (kinds of tasks, features of online environment, tutor's or teacher's strategy, etc.). In addition, it is possible to identify other patterns or "emerging roles", with reference to these contextual aspects. It is also possible to adopt a "scripted roles" perspective, which focuses on how the collaborative learning process can be facilitated by structuring and prescribing role by the teachers to learners (Cesareni, Cacciamani & Fujita, 2016). Emerging and scripted roles can be analyzed in terms of SCF to study their development during all the online course. It would be interesting to study the impact of these roles, through the social network analysis (Mazzoni & Bertolasi, 2012), on the students participation. It is possible to use the Weighted Indegrees (number of ties that a member of the group receive from other students) as indicator of "social attractivennes" and Weighted Outdegrees (number of ties that a member of the group outcome toward other members) as indicator of "social influence" exercised through these roles towards other course members (Weimann, 1994; Cacciamani, 2017). These parameters can allow to analyze what contribution each student, assuming an emerging or a scripted role, can offer in the collaborative knowledge building activity and which level of social influence and attractivennes he or she has on the community.

Finally, CF4KB coding scheme can be used in order to create a user profile for the participants, according to the CF they perform. In fact, although collaborative online environments have been extensively studied as communities (Chen & Caropreso, 2004), there has been a growing interest to analyze individual differences in collaborative environments. Such a study will help to understand individual's differences in terms of the different kinds of CF performed.

Acknowledgments

Stefano Cacciamani designed the research project, worked on data collection and analysis and, in the present article, has written the section Conclusions.

Vittore Perrucci worked on data collection analysis and in the present article has written the section of Method and Results.

Ahmad Khanlari helped with statistical data analysis using SPSS and worked on the introduction. In the present study, he has written the introduction section.

All the authors contributed to the final version of the article through reciprocal supervision.

REFERENCES

Bereiter, C. (2002), *Education and Mind in the Knowledge Age*, Hillsdale, NJ: Erlbaum.

Bereiter, C., Scardamalia, M., Cassells, C., & Hewitt, J. (1997), *Postmodernism, Knowledge Building, and elementary science*, The Elementary School Journal, 97(4), 329-340.

Cacciamani S. (2017), *Experiential Learning and Knowledge Building in Higher Education: an Application of the Progressive Design Method,* Journal of e-Learning and Knowledge Society, 13 (1), 27-38.

Cesareni, D., Cacciamani, S., & Fujita, N. (2016), *Role taking and knowledge building in a blended university course,* International Journal of Computer-Supported Collaborative Learning, 11(1), 9-39 .

Cacciamani, S, Perrucci, V. e Khanlari, A. (2016), *Knowledge Building and Conversational Functions in online interactions: a coding scheme.* In: Looi, C. K., Polman, J. L., Cress, U., and Reimann, P. (Eds.), Transforming Learning, Empowering Learners: The International Conference of the Learning Sciences (ICLS) 2016, 1195-1196, Volume 2, Singapore: International Society of the Learning Sciences.ISBN:978-0-9903550-8-3. Available at: https://www.isls.org/icls/2016/

Chen, S., & Caropreso, E. (2004), *Influence of personality on online discussion,* Journal of Interactive Online Learning, 3(2), 1-17

Hare, A. P. (1994), *Types of roles in small groups: A bit of history and a current perspective*, Small Group Research, 25(3), 433–448.

Landis, J.R., & Koch, G.G. (1977), *The measurement of observer agreement for categorical data,* Biometrics, 33, 159-74.

Lipponen L., Hakkarainen K., & Paavola S. (2004), *Practices and orientations of CSCL* In: Strijbos J., Kirschner P. & Martens R.(eds.). What we know about CSCL, and implementing it in higher education, 31-50, Boston, MA: Kluwer Academic Publishers

Mazzoni, E., & Bertolasi, S. (2012), *La Social Network Analysis (SNA) applicata alle comunità virtuali per l'apprendimento: analisi strutturale delle interazioni* *all'interno dei Web forum,* Journal of e-Learning and Knowledge Society, 1(2), 243-257.

Paavola, S., & Hakkarainen, K. (2005), *The knowledge creation methaphor- an emergent epistemological approach to learning,* Science and Education, 14, 535-557.

Pena-Shaff, J, B.,& Nicholls, C. (2004), *Analyzing student interactions and meaning construction in computer bulletin board discussions,* Computers & Education 42, 243–265

Scardamalia, M., & Bereiter, C. (2003a), *Knowledge Building,* in: Encyclopedia of education (1370-1373). New York: Macmillan Reference, USA.

Scardamalia, M., & Bereiter, C. (2003b), *Knowledge building environments: Extending the limits of the possible in education and knowledge work.* In: Di Stefano A., Rudestam K.E., & Silverman R. (Eds.), Encyclopedia of distributed learning, 269-272, Thousand Oaks, CA: Sage Publications.

Scardamalia, M. & Bereiter, C. (2006), *Knowledge building: Theory, pedagogy, and technology.* In: Sawyer R. K (Ed.), The Cambridge handbook of the learning sciences, 97-115, Cambridge, Cambridge University Press.

Stahl, G. (2003), *Building collaborative knowing: Elements of a social theory of learning.* In: Strijbos, J.W., Kirschner P. & Martens R. (Eds.), What we know about CSCL in higher education, 53-85, Amsterdam, NL: Kluwer.

Strijbos, J. W., Martens, R. L., Prins, F. J., & Jochems, W. M. (2006), *Content analysis: What are they talking about?,* Computers & Education, 46(1), 29-48.

Strijbos, J. W. & Weinberger, A. (2010), *Emerging and scripted roles in computer-supported collaborative learning,* Computers in Human Behavior, 26, 491-494.

Weimann, G. (1994), *The Influential: People who influence people,* Albany, NY: State University of New York Press.

Wise, A. F., Saghafian, M., & Padmanabhan, P. (2012), Towards more precise design guidance: specifying and testing the functions of assigned student roles in online discussions, *Educational Technology Research and Development*, *60*(1), 55-82.

Je-LKS
PEER REVIEWED PAPERS
Vol. 14, n.1, 2018
ISSN: 1826-6223 | eISSN: 1971-8829
Journal of e-Learning and Knowledge Society
The Italian e-Learning Association Journal

MOBILE APPLICATIONS IN UNIVERSITY EDUCATION: THE CASE OF KENYA

Ronald Ojino[1]
Luisa Mich[2]

[1] Directorate of Computing and eLearning, The Co-operative University of Kenya, Kenya - ronojinx@gmail.com
[2] Department of Industrial Engineering, University of Trento, Italy
luisa.mich@unitn.it

Keywords: Apps, Higher education, Mobile learning, Mobile phones, e-learning.

The widespread adoption of mobile phones has brought an increasing interest in the development of mobile applications for higher education. In this paper we examine the use of mobile applications in university education, focusing on Kenya, a leading country in mobile services. The main goal is to investigate if university students are using or would like to use mobile phones and apps, in particular educational apps. Information gathered from the study gives an insight on which apps to adopt or implement in university education.

for citations:
Ojino R., Mich L. (2018), *Mobile Applications in University Education: the case of Kenya,* Journal of e-Learning and Knowledge Society, v.14, n.1, 109-123. ISSN: 1826-6223, e-ISSN:1971-8829
DOI: 10.20368/1971-8829/1369

1 Introduction

The mobile phone is a ubiquitous device that many cannot imagine living without. This device is a widespread form of personal technology that is used for making and receiving phone calls, sending and receiving text messages, video and audio capture, and basic editing. In addition, it is used as a personal organizer, and for accessing mobile applications (apps) that provide a myriad of services in healthcare, education, agriculture, finance, hospitality, governance and environmental services among others. In the education sector, mobile applications have been identified as one of the six technologies to be watched for higher education advancement (Horizon, 2012). More importantly for our study, 90% of online time is spent using mobile apps (Smartinsights, 2016).

In this paper we will investigate the use of mobile apps in university education, focusing on Kenya, a leading country in mobile adoption and services (Murray, 2015), as confirmed by data on mobile penetration and mobile apps development (CAK, 2016). The main goal is to investigate if university students are using or would like to use mobile phone apps, in particular educational apps. Information gathered from the study will give insight on which apps could be adopted or implemented in university education. An analysis of some innovative technologies which when used with mobile apps could leverage their impact is also proposed.

According to the Global Digital report, almost two-thirds of the world's population currently has a mobile phone, and more than half of this uses smartphones; in addition, over half of the world's web traffic currently comes from mobile phones (Wearesocial, 2017).

In Africa, while Internet user penetration is only 29% (compared to the 50% of the world average), mobile phone penetration reached 71% in January 2017. In Kenya, Internet users penetration was 78% in March 2017 (compared to 45% in 2016 and 7.5% in 2006); mobile user penetration was 88.1% in 2015 and reached 90% in 2017 (CAK, 2016).

As for the apps market, worldwide downloads grew by 15% from 2015 to 2016, and the time spent using apps grew by 25%. Google Play and the iOS App Store with 2,800,000 and 2,200,000 apps respectively in March 2017 lead the mobile apps market (App Annie, 2016). Available data shows that the most popular App store categories by share of apps market are games (25.04%), business (9.88%) and education (8.36%).

Focusing on Africa, 2.4 million direct jobs were created in Sub Saharan Africa by the mobile ecosystem in 2013, and this is expected to rise to 3.5 million by 2020 (GSMA, 2014). Furthermore, the mobile ecosystem has deepened democracy through citizen participation (HellStrom, 2010), enhanced social ties due to more frequent use (Shrum *et al*., 2011) and improved livelihoods, among others.

Nicknamed the Silicon Savannah, Kenya is a re-known hub for innovative mobile applications in Africa. Applications incubated and popularly used in Kenya include, MPESA, a money transfer system that uses SMS to facilitate payment of bills, sending and receiving cash, banking and purchasing of products; Ushahidi, a crowd-sourcing app for sharing crisis information; M-Farm, that informs farmers about current market prices, agricultural trends and offers them the ability to collaborate; MedAfrica that helps diagnose symptoms, offers prescriptions, authenticates counterfeit drugs and directs patients to the nearest hospital if all interventions fail[1].

Focusing on the education sector, this paper presents the results of a survey carried out across three Kenyan universities in 2015. The main objective of the study was to identify the mobile apps used for educational purposes by students in Kenyan universities, and to investigate the apps and services considered useful by students. From the perspective of universities and mobile application developers, the goal of this paper is to provide insights about the potential demand of educational mobile apps for use in Kenyan universities, and in turn other universities across Africa.

The rest of the paper is organized as follows. Section 2 introduces some classifications for mobile apps in education. The survey and the main results are presented in the third section. Section 4 introduces some of the most innovative technologies which when used with apps could leverage their impact. Finally, the last section gives the conclusion, and future work.

2 Uses of mobile phones in university education

In university education, the ubiquitous mobile phones present a huge potential because of the large number of activities and tasks they can support (Valk *et al*., 2010). The USA ranks first in the field of mobile learning probably due to the fact that it is one of the best states in the world in terms of technology (Soykan & Uzunboylu, 2015). In Japan, m-learning already has a rich and vibrant history and in 2005, practically 100 percent of college students and working adults in Japan owned a mobile phone (Kato & Ricci, 2006). However, a systematic review by Alioon and Delialioglu (2015) revealed that m-learning projects have been considered more in developing countries than developed ones as they are cost effective. Besides, many m-learning projects have been applied dominantly for K12 environments rather than higher education and the most frequently used approaches for implementing m-learning are mobile applications, followed by SMS and mobile game projects respectively, whereas smartphones are the most common devices in terms of m-learning. The down-

[1] www.safaricom.co.ke; www.ushahidi.com; www.mfarm.co.ke; www.medafrica.org

sides to mobile learning that may impede its adoption include: actual feasibility depends on the interest and diligence of learners (Kukulska-Hulme, 2005); wireless technology may require universities to impart successful degrees to the same caliber of students, if mobile learning is to be included as a mainstream education platform (Ally, 2009); security issues and designing a common user interface itself is a challenge (Alrasheedi & Capretz, 2015). In Switzerland, data collected from 2 universities in Ticino showed that only 17.3% of students consider mobiles important tools for learning, and while 42% of students use mobile phones for learning, 3 out of 4 use them to interact with their peers (Rapetti *et al*., 2011). Many instructors believe that students use mobile phones for socializing purposes when they reported that they were doing study-related tasks (Pollara, 2011). As much as many learners are receptive of m-learning, instructors are skeptical of the idea and are slow to adopt it (Alrasheedi & Capretz, 2015). Focusing on Africa, a high percentage (over 90%) of university students in Nigeria's Kwara State have a positive attitude towards mobile learning with many concurring that if adapted it would enhance learning and over 85% being ready to adopt it if introduced (Adegbija & Bola, 2015).

Few researchers have attempted to categorize mobile apps use in education. Laurillard (2002) developed a conversational framework for the effective use of learning technologies, which can be applied in a range of subjects. Kole (2009) developed FRAME (Framework for the Rational Analysis of Mobile Education) which classified mobile learning according to three characteristics, namely: the device, the learner and the social environment, thereby highlighting the social, personal and technical aspects which can help determine the effectiveness of mobile learning. FRAME is a comprehensive model that is useful in planning and designing mobile learning environments because it suggests a checklist of questions that can guide the development process. The framework by Park (2011) has four types of mobile learning which are: *high transactional distance socialized m-learning* where learners are involved in group learning and have more communication space with their instructor; *high transactional distance individualized m-learning* in which individual learners receive tightly structured content and control their learning process to master it; *low transactional distance socialized m-learning* where learners have less psychological and communication space with the instructor as well as loosely structured instruction and work together in a group and *low transactional distance individualized m-learning* in which there is less psychological space between instructor and learner, is loosely structured and the instructor leads the learning. Patten *et al*. (2006) developed a functional framework for categorizing handheld educational applications that views the mobile learning design space in terms of application functions and pedagogical underpinning (Figure 1).

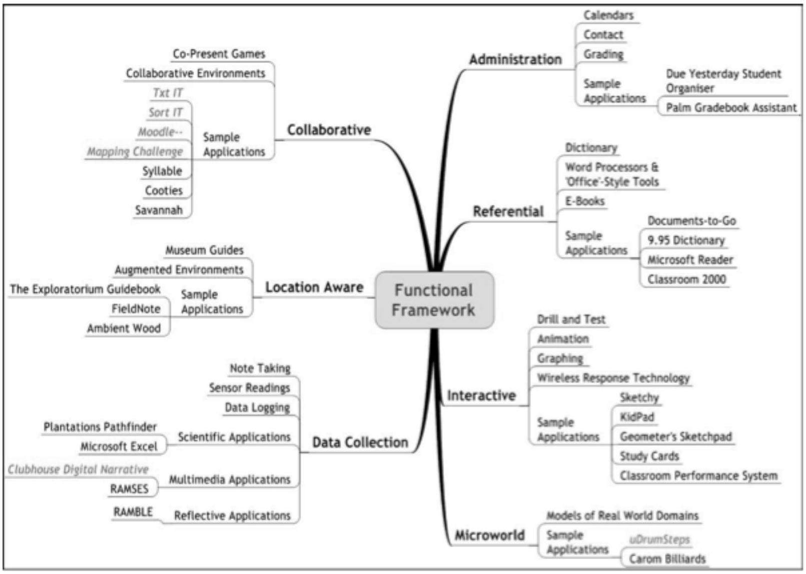

Fig. 1 - Functional framework (Patten *et al.*, 2006)

The seven categories in the model are: *administrative apps* which focus on informative storage and retrieval while replicating available tools on traditional platforms; *referential apps* that allow accessing of content at learning places; *collaborative apps* that encourage knowledge sharing in the learner's physical context and enhance collaborative learning principles; *location aware apps* which contextualize learning activities by facilitating appropriate interaction among learners and their environment; *data collection apps* used to record reflective, scientific and multimedia data about the environment; *interactive apps* that focus on content delivery and information management via a 'response and feedback' approach; and *microworlds* that allow learners to construct their own knowledge through experimentation in constrained models of real world domains. These categories have been used to identify and classify apps for universities (Table 1), and then to extract those to include in the questionnaire used for the survey.

Table 1
CLASSIFICATION OF APPS BASED ON THE FUNCTIONAL FRAMEWORK

Class of applications	Examples of mobile app functions
1.Administrative apps	Attending classes (virtual classroom) Authentication when entering the campus Booking rooms of residence Course registration Delivery of lectures and other course materials Fees payment Making notes during class Monitoring student progress Mind-mapping (for mapping out ideas) Preparation of lecture notes Provision of help desk information Recording of class attendance Relaying of campus news and events Time tabling Uploading of assignments Uploading and dissemination of exam results Voting and polling processes
2.Referential apps	Dictionary eReaders Language translators Mobile tactile braille Provision of access to research databases Reminders Searching library catalogues Sign language interpreter Sign language learner Text-to-speech app
3.Interactive apps	Answering questions in class Creating flash cards Educational games Lab experiments simulators Unit (subject) focused apps Unit (subject) examination/quiz
4.Microworld	Creating apps in computing courses Creating podcasts Creating videos Making presentations
5.Data collection	Analysis of data collected during research Data collection
6.Location aware	Provision of self-guided campus tours Virtual maps
7.Collaborative	Collaborative writing Facilitation of discussions among classmates Facilitation of interactivity between students and lecturers Inviting guest lecturers

3 The study

The goal of the study was to investigate the potential demand of mobile apps within Kenyan universities. It was conducted at the University of Nairobi, The Co-operative University College of Kenya and Great Lakes University of Kenya. These universities were chosen because they are representative of the higher education institutions landscape in Kenya, which include established public universities, university colleges and private universities. The study was based on a structured questionnaire that was completed online and on hard copy. Trained research assistants interviewed students at the three universities randomly selected from their class lists, and entered the responses on the paper-based questionnaires. The population consisted of both undergraduate and postgraduate students.

An on-line version of the questionnaire was offered to students who were unable to meet with the researchers. In this way it was also possible to gather informal comments and remarks. The questionnaire had 16 questions. The goal of the first 5 questions was to investigate the ownership and use of mobile phones. Questions 6 to 14 examined different topics related to the apps; apps installed on the mobile phone (number and kinds); types of mobile apps used for academic purposes and their impact on learning processes and results; mobile apps hosted by the university if any; usefulness of a set of relevant mobile apps a university could offer to its students and suggestions for any other apps that could be offered by the university. The last two questions sought to know the respondents' university and the age range. The survey was run in autumn 2015 and sampled a total of 134 respondents; over 99% were aged 35 and below. A pre-test of the questionnaire was carried out at The Co-operative University College of Kenya's faculty of commerce.

3.1 Results

The first result of the study confirmed that a huge percentage of students (95.6%) who participated in the survey owned a mobile phone, a percentage higher than the 82% of all the population in Kenya (Wearesocial, 2017). 89.6% of the respondents owned a smartphone, compared to the 44% for all the population (Google Barometer, 2016). In the case of Kenyan students, the percentage is also higher than that measured in the US, 86% (Pearson, 2015). This is possibly due to the affordability of mobile phones in Kenya that cost approximately 10 $ for the cheapest feature phone and 50 $ for the cheapest smart phone. Only 35.2% of the respondents had owned the mobile phone for more than 2 years (41% for less than 1 year), showing that most of the students are in possession of new generation mobile phones.

Answers to the question on how long they can do without a mobile phone exposed that 77.7% of the respondents could not spend more than one day without accessing their mobile phones: 39.6% could stay without their mobile phone for less than 1 hour, while 26.5% for between 1 to 5 hours maximum. For most of the students using a mobile phone had become a habit and they felt the need to stay connected at all times (41.8%). The findings are consistent with those available in (Wearesocial, 2017; Pearson, 2015). With regard to the frequency of mobile phone usage, most students (83.5%) used their mobile phones between 1 to 10 times a day.

Another question investigated the type of default applications installed on respondents' mobile phones. The findings show that the majority of students had installed messaging (text) apps (83.6%), calculators (76.9%), browsers (76.1%), alarms (74.6%), recorders (70.1%), banking/shopping apps (69.4%), a percentage higher than in most of the other countries (Statista, 2014), convertor (60.4%), organizers (56.0%) and weather forecasting apps (44.0%). As for the number of non-default apps installed in the students' mobile phones, the majority of them (78.4%) had installed between 2 to 15 non–default apps. Of these, it was reported that games (69.4%), and social networking apps (61.9%) were the most popular; e-reader and instant messaging are used by 35.1% of the students; followed by news (29.9%) and video streaming (29.1%) which depict that the students mainly use their mobile phones for communication and entertainment. Gambling apps are at 14.9%; travel, television and e-ticketing stood at about 10%. Focusing on educational apps, only 3.7% of the students did not use any. Among the most popular academic uses of mobile applications were: accessing a dictionary (63.4%), accessing course materials (61.9%), and registering for courses (61.2%); about half of the students use apps to download research publications (53.7%), to download assignments (51.5%); a smaller percentage used it for undertaking quizzes or exams (20.1%) and playing educational games (15.7%) (Table 2).

Table 2
ACADEMIC PURPOSES OF USING MOBILE PHONE APPLICATIONS

Purpose	Number	%
Use the dictionary	85	63.4
Access course materials	83	61.9
Register for courses	82	61.2
Download research publications	72	53.7
Download assignments	69	51.5
Take lecture notes	48	35.8
Hold class discussions	41	30.6

Purpose	Number	%
Reserve library resources	28	20.9
Use language translators	28	20.9
Undertake quizzes/examinations	27	20.1
Play educational games	21	15.7
Do not use for academic purpose	5	3.7
Others	2	1.5

Respondents were then asked to indicate the type of the educational functions they would have on mobile phone. A majority of students gave the following rankings: 56.7% of the students desire apps for receiving/submitting assignments, and for class timetables; 53% desire to receive notifications of exams grades. More than one third of them desire to receive campus news and to access the campus e-library (Figure 2).

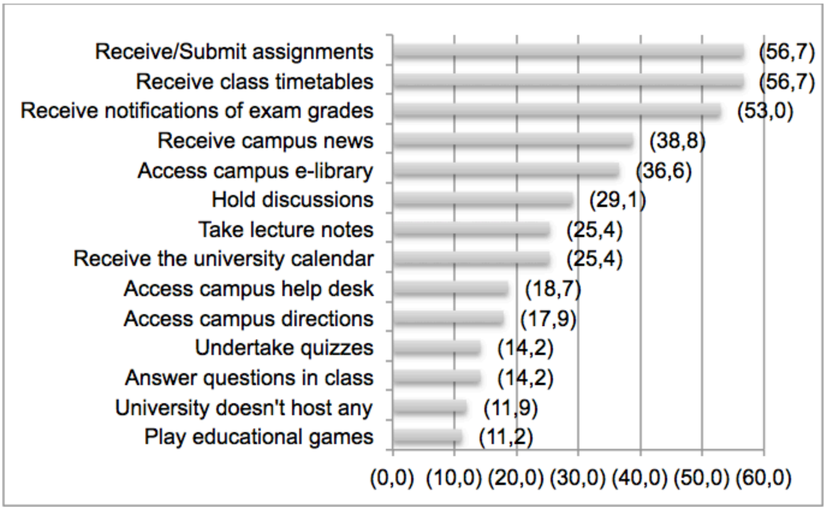

Fig. 2 - Educational mobile application functions desired by students (%)

Data regarding the usefulness of educational mobile apps showed that for students the most useful apps are those for receiving quiz and exam results (82.1%), accessing campus e-library (80.6%), receiving and submitting assignments (76/1%), and holding discussions (73.1%) (Table 3).

Table 3
PERCEIVED USEFULNESS OF MOBILE APPLICATIONS (%)

Apps	Not useful	Neutral	Useful	Missing
Accessing campus e-library	10.4	5.2	80.6	3.7
Attending virtual classes	21.6	14.9	59.0	4.5
Creating videos and podcasts	25.4	16.4	52.2	6.0
Holding discussions	14.2	8.2	73.1	4.5
Mapping out ideas	13.4	11.9	64.2	10.4
Playing educational games	25.4	24.6	44.0	6.0
Receiving and submitting assignments	9.7	6.7	76.1	7.5
Receiving quiz and exam results	5.2	8.2	82.1	4.5
Taking lecture notes	14.2	12.7	69.4	3.7
Undertaking quizzes and exams	18.7	11.9	64.2	5.2

3.2 Discussion

Students sampled in the study were found to spend an ample amount of time engaged in them. Most of the students had default apps, as messaging and social networking apps, installed on their phones, which is consistent with findings from previous studies by (Zulkefly & Baharudin, 2009; Lie, 2004; Maddel; Muncer, 2004; and more recent data reported by Google and Wearesocial). This is probably because the cost of communication via SMS and social networking tools is lower than voice calls, and students have limited financial resources.

As regard educational apps, many students desire to have apps that would help them submit and receive assignments as well as timetables and campus news. This is consistent with the trend in which people want easy and timely access to not only the information on the network, but also to tools, resources, and up-to-the-moment analysis and commentary (Horizon, 2012).

Most of the students perceive mobile apps for holding discussions to be useful also in an educational context. This is probably because they are already accustomed to apps for social networking communication, where they can easily set up groups and carry on with discussions. This trend could also be a result of the increased uptake of social media as evidenced by the rise of mobile social media use in Africa by nearly 50% during the year 2016 (Wearesocial, 2017).

Findings from the study highlight that a high number of students would like to access e-library services online, whereas below half of the ones sampled do have such service. The interest in accessing e-library services is possibly due to the rise in internet and data subscriptions (CAK, 2016) as well as the wide range of information that students could exploit when carrying out assignments and research.

Other results to be highlighted are those related to educational games: many students did not play and did not desire to have any educational games apps. This finding is contrary to the postulation by Brigham (2015) that many people are attracted to gamification as it enhances participation and engagement. A reason for this contradiction could be that this concept has not been applied across local institutions in Kenya and thus educational game apps cannot be perceived as useful by students.

4 Technologies for leveraging educational apps

Novel technologies are blurring the line between education and leisure and can be used with mobile phones in education to develop new and innovative mobile apps. For example, all the universities surveyed offer e-learning, though the mode of delivery is rudimentary since the course content is offered mainly through PDFs. The universities need to make their e-learning systems more interactive through educational apps. Technologies that can be applied with apps to address that challenge and to support functions desired and considered useful by students (Figure 2 and Table 3) include: augmented reality (AR), Internet of Things (IoT), mobile learning analytics and game based learning.

Augmented Reality by inserting virtual information into the real world through apps allows enhancement of a user's perception to learn about and annotate his environment (FitzGerald *et al*., 2013), promoting engagement and motivation (Klopfer & Squire, 2008; Luckin & Stanton Fraser, 2011; Kesim & Ozarslan, 2012).

In the IoT, the Internet connecting physical things, mobile apps can be used as control devices or actuators (Want *et al*., 2015) and let the students take an active part in many activities (Bandyopadhyay & Sen, 2011). For example, IoT apps have been applied to learn about cultural attractions (Chianese & Piccialli, 2014), to record students' class attendance automatically, to access smart packing spaces (Nie, 2013), to locate places within campus and to personalize students' apartments.

Mobile learning analytics refers to the collection, analysis and reporting of the data of mobile learners, which can be collected from the mobile interactions between learners, mobile devices and available learning material (Aljohani & Davis, 2012). The purpose is to observe and understand learning behavior to enable appropriate intervention (Educause, 2011). The final feedback gives not only the overview of correct solutions, but also a detailed analysis of typical errors (Ebner *et al*., 2014).

Game based learning, or Gamification uses game features - elements, mechanics, frameworks, aesthetics, thinking, metaphors - into non-game settings and support more interactivity while learning (Faiella & Ricciardi, 2015; Notari, *et al*. 2016; Prensky & Prensky, 2007). The *Serious Games* movement has focused on uniting significant educational content with play (Horizon, 2012). Examples of gamified mobile apps to enhance learning are Duolingo, a free language-learning app where users are provided instant feedback and gain 'experience points' (Brigham, 2015); SCVNGR is a location-based mobile gamification app with customizable treks and challenges, so that students can discover more about the school and the campus in the form of a scavenger hunt (Keller, 2011); Fantasy Geopolitics motivates students to learn more about their countries by news reading (Magdaleno, 2014). Library Quest engages users by asking them to input alphanumeric codes or to scan QR codes displayed in the library building to encourage them to explore the library building and to make them aware of various library services (Felker, 2013).

Conclusions

For any university that is innovative and continuously exploring new strategies for education, ignoring the potential of mobile apps is generally detrimental to progress, and it would imply refusal to adapt in a continuously changing world. The study based in Kenya, throws some light on the apps a University could offer to its students. Almost all of them have a focused scope and are meant for daily use. These results are possibly useful for educational apps anywhere in the world. According to the results of the survey, there is a large potential arising for educational institutions to exploit mobile apps in enhancing learning; but also administration and other activities could benefit from students behaviors and needs in relation to mobile technologies and apps. For example, across the Kenyan Universities, students currently utilize commercial applications such as MPesa, useful also for the administration for purchasing and payments (in tandem with the high volume of mobile commerce transactions on a rising trend in Kenya (CAK, 2016)).

From the possible adoptions and uses of mobile apps in education, it is notable that most of the initiatives are prevalent in the developed world (http://www.unesco.org) possibly due to sustainability issues (Traxler & Leach 2006), while data on mobile and apps usage show high percentages in African countries; highlighting a huge potential for demand and use. Education institutions and universities especially in the developing world should consider incorporating mobile apps in learning as this would optimize the ways in which students learn by opening up new education doors and improving efficiency in carrying out of administrative duties. It is also important to note that if mobile apps usage in universities lack directed learning activities, they may be detrimental to the overall learning process and this study offers preliminary insights on the kind of apps that could be adopted and developed to support university processes.

As for the technologies illustrated in section 4 that could enhance the performances and the scope of apps, Kenyan universities are lagging behind in comparison to their competitors in the developed nations. Of the three universities surveyed, only the University of Nairobi has a laboratory that offers training on IoT in which startups work on a range of IoT projects. This situation is probably due to the lack of funds limiting the investment of the universities in the latest technologies. Kenyan universities could then collaborate more with world re-known universities and be more innovative in generating funds to support the integration of these technologies. Future work include an analysis of their adoption to implement innovative educational apps addressing the needs highlighted by students. Another area that needed to be investigated is that of mobile apps and MOOCS (Massively Open Online Courses). None of the universities in the survey offers or has adapted online courses offered through MOOCs such as Udemy, edX or Coursera. Through MOOCs there lies a great potential which if exploited would imply a better global presence of Kenyan universities, as their courses would be available worldwide, faculty could have a wider audience for selling course materials, and the institutions could make savings on staffing needs, among other benefits.

REFERENCES

Adegbija M.V. & Bola O.O. (2015), Perception of undergraduates on the adoption of mobile technologies for learning in selected universities in Kwara state, Nigeria. *Procedia-social and behavioral sciences*, 176, 352-356.

Alioon, Y., & Delialioglu, O. (2015), A Frame for the literature on m-learning, *Procedia-social and behavioral sciences*, 182, 127-135.

Aljohani N.R. & Davis H. (2012), Significance of learning analytics in enhancing the mobile and pervasive learning environments, *Proc. NGMAST*, pp. 70-74. IEEE.

Aljohani N.R. & Davis H.C. (2013), Learning analytics and formative assessment to provide immediate detailed feedback using a student centered mobile dashboard, *Proc. NGMAST,* pp. 262-267. IEEE.

Ally M. (2009), *Mobile learning: transforming the delivery of education and training,* Edmonton, Alberta, Canada: Athabasca University Press.

Alrasheedi M. & Capretz L.F. (2015), An empirical study of critical success factors of mobile learning platform from the perspective of instructors, *Procedia-social and behavioral sciences*, 176, 211-219.

Andersson A. (2007), Beyond student and technology: Seven pieces to complete the e-learning Jigsaw puzzle in developing countries, *Proc. IRIS30*.

App Annie (2016), Retrospective. Mobile's continued momentum. http://go.appannie.com/app-annie-2016-retrospective-thank-you.html?aliId=181259012

Brigham T.J. (2015), An introduction to gamification: adding game elements for engagement, *Medical reference services quarterly*.

Chianese A. & Piccialli F. (2014), Designing a smart museum: When cultural heritage joins IoT, *Proc. NGMAST*, pp. 300-306. IEEE.

CAK - Communications Authority of Kenya (2016). http://ca.go.ke/index.php/component/ content/category/99-research-statisticsz

Ebner M., Schön M. & Neuhold B. (2014), Learning analytics in basic math education– first results from the field, *eLearning papers*, 36, 24-27.

Felker K. (2013), Library quest: Developing a mobile game app for a library, *ACRL TechConnect Blog*, http://acrl.ala.org/techconnect/?p=3783.

FitzGerald E., Ferguson R., Adams A., Gaved M., Mor Y. & Thomas R. (2013), Augmented reality and mobile learning: the state of the art, *Int. Jour. of mobile and blended learning*, 5(4) 43-58.

Google Barometer (2016), *Consumer barometer*, www.consumerbarometer.com/en/ trending/?countryCode=KE&category=TRN-NOFILTER-ALL.

GSMA (2014), The mobile economy - Sub Saharan Africa 2014. *Intelligence report*.

Hellström J. (2008), Mobile phones for good governance–challenges and way forward, *Upgraid*, www.w3.org/2008/10/MW4D_WS/papers/hellstrom_ gov.pdf.

Horizon (2012), *NMC Horizon report*, Higher education edition, www.nmc.org/pdf/2012-horizon-report-HE.pdf.

Kato M.K. & Ricci V. (2006), Mobile learning in Japan: Why the future has already arrived in Asia, *The e-learning guild's learning solutions*, pp. 1-11.

Keller J. (2011), Smartphone game turns college tours, *Wired campus*, http://chronicle.com/blogs/wiredcampus/mobile-game-turns-college-tours-and-orientations-into-scavenger-hunts/33114

Kesim M. & Ozarslan Y. (2012), Augmented reality in education: current technologies and the potential for education, *Procedia-social and behavioral sciences*, 47, 297-302.

Kukulska-Hulme A. (2005), Introduction. In A. Kukulska-Hulme, & J. Traxler, *Mobile learning: a handbook for educators and trainers* (pp. 1- 6). New York: Roultedge.

Laurillard D. (2002), *Rethinking university teaching, A conversational framework for the effective use of learning technologies*. London: Routledge.

Magdaleno A. (2014), How a high school teacher is 'gamifying' world news, *Mashable*, http://mashable.com/2014/02/02/high-school-fantasy-geopolitics.

Murray S. (2015), *Africa: Huge global investment ploughed into Africa's telecoms industry*, http://allafrica.com/stories/201506222526.html.

Notari M.P., Hielscher M. & King M. (2016), Educational apps ontology, *Mobile learning design* (pp. 83-96). Springer Singapore.

Nie X. (2013). Constructing smart campus based on the cloud computing platform and the internet of things, *Proc. ICCSEE 2013*.

Park Y. (2011). A pedagogical framework for mobile learning: Categorizing educational applications of mobile technologies into four types, *The international review of research in open and distributed learning*, 12(2) 78-102.

Patten B., Sánchez I.A. & Tangney B. (2006), Designing collaborative, constructionist and contextual applications for handheld devices, *Computers & education*, 46(3)

294-308.
Pearson (2015), *Student mobile device survey: college students*, www.pearsoned.com/wp-content/uploads/2015-Pearson-Student-Mobile-Device-Survey-College.pdf.
Pollara P. (2011), *Mobile learning in higher education: A glimpse and a comparison of student and faculty readiness, attitudes and perceptions.* Baton Rouge: Louisiana Univ.
Prensky M. (2007), *Digital game-based learning*. St. Paul, MN: Paragon house.
Rapetti E., Picco-Schwendener A. & Vannini S. (2011), Is mobile learning a resource in higher education? Data evidence from an empirical research in Ticino (Switzerland). *Journal of e-learning and knowledge society*, 7(2), 47-57.
Ring G. (2001), Case study: Combining Web and WAP to deliver e-learning. *Learning circuits*.
Shrum W. *et al*. (2011), Mobile phones and core network growth in Kenya: Strengthening weak ties, *Social science research* 40(2), 614-625.
Smartinsights (2016), *Mobile marketing analytics*, www.smartinsights.com/mobile-marketing/mobile-marketing-analytics/mobile-marketing-statistics/attachment/percent-time-spent-on-mobile-apps-2016.
Soykan E. & Uzunboylu H. (2015), The review of published articles on mobile learning area in EBSCO database, *Procedia-social and behavioral sciences*, 182, 710-717.
Statista (2014), *Usage of mobile banking apps worldwide in 2014*, www.statista.com/statistics/468943/usage-of-mobile-banking-apps-worldwide-by-country.
Traxler J. & Leach J. (2006), Innovative and sustainable mobile learning in Africa, *Proc. WMTE*. IEEE
Valk J.H., Rashid A.T. & Elder L. (2010), Using mobile phones to improve educational outcomes: An analysis of evidence from Asia. *The Int. review of research in open and distributed learning*, 11(1), 117-140.
Want R., Schilit B.N. & Jenson S. (2015), Enabling the Internet of Things. *IEEE Computer*, 48(1), 28-35.
Weraresocial (2017). *Digital in 2017: Global overview*, https://wearesocial.com/blog/2017/01/digital-in-2017-global-overview.

Je-LKS
Journal of e-Learning and Knowledge Society
The Italian e-Learning Association Journal
PEER REVIEWED PAPERS
Vol. 14, n.1, 2018
ISSN: 1826-6223 | eISSN: 1971-8829

ANALYSING ACCESSIBILITY, USABILITY AND READABILITY OF WEB-BASED LEARNING MATERIALS – CASE STUDY OF E-LEARNING PORTALS IN SLOVENIA

Marko Radovan
Mojca Perdih

Faculty of Arts, University of Ljubljana
marko.radovan@ff.uni-lj.si, mojca.perdih@gmail.com

Keywords: Web-based learning materials, Dyslexia, Accessibility, Usability, Readability.

In this article, by utilizing the guidelines available from literature, we attempted to compile a checklist that would identify the extent to which web-based learning materials in Slovenia properly address the needs of people with dyslexia. The focus of our research was the evaluation of the suitability of the design of web-based learning materials and not their pedagogical aspects, since design itself is one of the key factors that predominantly affect the accessibility of web-based learning materials for dyslexic learners. The results of our study showed that the developers of web-based learning materials were generally well aware of these criteria (accessibility, usability, readability); but some of the key ones remained overlooked. The study pinpoints accessibility as the weakest point of the examined web-based learning materials. The most plausible reasons for overlooking these criteria originate in the weak understanding of the needs of these particular users in addition to the weak comprehension of the established guidelines on the

for citations:
Radovan M., Perdih M. (2018), *Analysing Accessibility, Usability and Readability of Web-based Learning Materials – Case study of e-learning portals in Slovenia,* Journal of e-Learning and Knowledge Society, v.14, n.1, 125-136. ISSN: 1826-6223, e-ISSN:1971-8829
DOI: 10.20368/1971-8829/1389

accessibility of web-based learning materials. As a potential solution, we suggested, apart from using the checklist as an indicator, increasing attention among the developers of web-based learning materials when considering accessibility issues as well as a wider integration of the concerned user population in the evaluation of web-based learning materials.

1 Introduction

Today's society perceives knowledge as a source and factor of a good-quality life. In the new social and technological settings, traditional educational concepts and methods no longer suffice, as they cannot cope with the increasing growth of knowledge, which also quickly becomes obsolete. The individual should have the opportunity to learn in all ages. Hence, traditional educational systems should become more open, flexible, adapted to individual learning objectives, individual needs, and interests. This is where e-learning comes very much to the foreground, because it supports new perspectives and possibilities of acquiring and creating knowledge (Agrusti, 2013; Dinevski & Radovan, 2013; Fee, 2009). One of its key characteristics is flexibility, which enables easier access to information for everyone. Not only people with dyslexia, but also all learning individuals can use these materials to gain better quality and friendlier access to educational content – as long as the materials are designed appropriately and allow for adaptation to individual needs.

Definitions vary mainly in how they perceive dyslexia (Camp & Aldridge, 2007; Doyle, 2002). Some authors emphasize its neurological factors, others educational characteristics, others still stress cognitive factors. Therefore, it is always important to take account of context when defining dyslexia, which will subsequently enable us to find the most suitable adaptations for the student's learning (Reid, 2009). The common ground all the theories share is the view of dyslexia as a developmental phenomenon, affecting the individual for life, with its main characteristic being difficulties in acquiring literacy skills. We should also highlight that there are numerous causes for dyslexia, including hereditary ones.

As for the learning of people with dyslexia, it is especially important to be aware that the consequences of dyslexia do are not compatible with the usual teaching methods. When tackling these people's different learning abilities, we can make great use of specific treatment, information and communications technology (ICT) and specific teaching adaptations (e.g. methods, techniques). Here, the suited approach of preparing learning materials is of great importance learners' satisfaction and its use (Mažgon, Šebart, & Štefanc, 2015).

2 The use of ICT in education and dyslectic learners

ICT can be of great help to people with dyslexia, but we should realize that it could not completely replace structured teaching and learning. This is also due to the constant appearance of new technologies and their relentless changes, which means that today technology is difficult to follow. Modern ICT is changing the educational process and educational content. This is not only true of teaching the content; it is providing learning individuals with a mass of resources, which are no longer only passive in form, as they include other media and interactive forms.

The wide variety of materials available on the market includes self-study online courses for individuals to learn as well as countless study notes accessible online (Smythe & Draffan, 2005). It is vital that these materials should be prepared in an effective, structured, organized manner and presented suitably, thus enabling users to acquire, understand, and process information as easily as possible. This is even more critical if users include people with dyslexia.

E-learning is not limited to merely multimedia materials and the Internet; it also involves assistive technology intended for users with special needs. For instance, when working with online materials, reading from the computer, listening to instructions or making notes, those with dyslexia frequently employ software such as text-to-speech software and digital voice recorders. Consequently, it is important for web-based learning materials to be developed to support the use of various software.

3 Guidelines for the development of web-based learning materials

As described previously (Radovan & Perdih, 2016), we attempted at developing a checklist to help us evaluate the adaptation of web-based learning materials to the needs of people with dyslexia. The checklist is based on three criteria: (1) **accessibility**, (2) **usability** and (3) **readability**. When drawing up the guidelines for each individual category, we relied on Rainger (2003), who puts forward practical recommendations for developing web-based learning materials for users with dyslexia, and on the book "Dyslexia in the Digital Age" by Smythe (2010), who is one of the leading authors in the field.

Our guidelines are therefore aimed at three sets of criteria that we think are essential for people with dyslexia:

- Accessibility. The accessibility of web-based learning materials is assessed according to whether a user with dyslexia can access information or not, regardless of how easy the materials are to use. They include: (1) enabling access to materials via assistive technology; (2) the use of illustrations, diagrams, flow-charts and photographs can enhance the accessibility of web-based learning materials; (3) enabling textual description of visual content etc.

- Usability. Learning content may be accessible, but if it is not embedded in a user-friendly environment, it will not lead to optimal educational/learning experience. As already stated, usability means how easily and how quickly we can learn from web-based learning materials. A possibility of assessing usability is counting the navigational errors that students make (their frequency and severity). The areas of usability we focus on are typography (typeface, font size, leading/line spacing, justification), text, background colour, and navigation.
- Readability. This is overlooked aspect of e-material design. It refers to the ease of understanding a text in terms of the vocabulary and grammar used. There are simple criteria for Internet material "readability", which are also important for dyslexic users, such as: (1) sentence length should be between 15 and 20 words; (2) explicit information structure (e.g. at the beginning, the learning/educational objectives, expectations, etc. should be emphasized); (3) instructions should be given clearly and without lengthy explanations; (4) use of bullet points or numbering where appropriate, etc.

3.1 Purpose of the study

Our research problem addressed the question about how web-based learning materials were adapted to people with dyslexia and we attempted to provide recommendations on how to improve them. The research study included freely accessible web-based learning materials for adults' independent learning in Slovenia. We selected a number of secondary school and higher education web-based learning materials and web-based learning materials for non-formal learning from various online portals offering online education.

Using a checklist, we evaluated how suitable the web-based learning materials were from the aspects of accessibility, usability, and readability. Therefore, our research questions referred to how web-based learning materials were adapted to people with dyslexia and to what degree they take account of the guidelines regarding each individual criterion. We focused on how the guidelines were applied in the areas of the accessibility, usability, and readability of web-based learning materials.

We asked the following research questions:

- Are the web-based learning materials suitable for use by people with dyslexia?
- How well applied are the guidelines in individual areas (i.e. accessibility, usability and readability)?

4 Methodology

Our sample consists of Slovenian online portals that publish web-based learning materials. The portals were searched for on Google, using the search string "web-based learning materials". When selecting the web-based learning materials there occurred the problem of their accessibility as some portals offer their web-based learning materials commercially, which means they are not freely accessible. Consequently, we opted for six portals with freely accessibly web-based learning materials that are designed for self-directed learning. In the next step, we randomly selected seven web-based learning materials from each web-portal, so we ended up with 49 web-based learning materials to evaluate.

Since we could find no adequate measuring instrument to assess the suitability of online materials for use by people with dyslexia, we attempted at developing a checklist to assist in answering the research questions. It is an evaluation checklist used to help assess the suitability of web-based learning materials regarding the needs of people with dyslexia (see Radovan & Perdih, 2016). The checklist covers 3 areas and 47 items. The first area is *accessibility*, containing 20 items, the second is *usability*, containing 18 items, and the third area is *readability*, containing 13 items. Each question had three possible replies: "Yes", "No", "Not possible to assess (NA)". We evaluated the online portals with web-based learning materials according to the share of the recommendations they apply in each area.

5 Results and discussion

5.1 General suitability of web-based learning materials for people with dyslexia

The first research question examined whether the web-based learning materials were sufficiently adapted for use by people with dyslexia. Since scales are not comparable, we standardised them to a unified scale from zero to 100.

Table 1 shows the descriptive analysis of the summary scales, with the values ranging from zero to 100 points, and a total of n = 49 analysed web-based learning materials in the sample. The higher the scale value, the more the web-based learning materials adhered to the guidelines in our checklist. We selected the middle of the scale as the test value (i.e. M = 50), with the values above M = 50 denoting adequate suitability and the values below M = 50 suggesting poor suitability.

Table 1
DESCRIPTIVE ANALYSIS OF THE SUMMARY SCALES AND THE UNIFIED SCALE

Accessibility Indicators	N	Nitems	Min	Max.	M	SD
Usability	49	20	60	93	84,21	10,51
Readability	49	18	50	100	76,36	12,42
Accessibility	49	13	25	70	51,53	12,79
Total	49	47	45	79	68,30	8,99

As the table above indicates, the web-based learning materials demonstrated different suitability according to different criteria. On average, they were the most suitable with regard to usability (M = 84.21 on the scale from zero to 100). On average, they were the least suitable with regard to accessibility (M = 51.53). The aspect of readability tended to be somewhere between the two (M = 76.36). It became apparent, that according to the unified scale, existing web-based learning materials reveal a generally satisfactory suitability for people with dyslexia, which means that their compilers largely take account of the criteria which are important to people with dyslexia.

A positive surprise was a good score in the area of usability, which mostly refers to the navigation, structure, and form of online contents. Usability may be a key factor in cognitive overload, which is likely to reduce learning effectiveness (Dunn, 2003). This is particularly significant for people with dyslexia, who sometimes have problems with short-term memory, which means that they may easily forget where in the material they find themselves, what they should click, and what sections of the material they have already studied.

Compared to the other two criteria in our sample, the criterion of accessibility scored slightly worse. On the one hand, accessibility requires enabling the user to access information in different formats (audio, video, pictures) and, on the other hand, it has to provide the user with more control over the representation of information (bigger font sizes, customising background colours, control over multimedia playing, etc.). Providing content in only one form and disabling customisation can have a considerable impact especially on people with dyslexia. It has been proven that these individuals can read more accurately if they are able to customize colours and font sizes (McCarthy & Swierenga, 2010), and they will remember the learning content more easily if it is supported by illustrations, diagrams, audio and video recordings (Burt, 2004), which they can see/hear several times.

In short, a general assessment suggests quite a positive result. However, in order to be able to interpret it we need to look into each individual criterion separately and establish what the guidelines with the lowest scores are and whether they are the ones with the strongest influence on people with dyslexia.

5.2 Applying the guidelines in the areas of the accessibility, usability and readability of online materials

5.2.1 Accessibility

The table below presents the shares of the application of each of the accessibility guidelines. The answers are given in descending order from the largest to the smallest share. Our analysis reveals that as many as nine guidelines on adapting web-based learning materials to people with dyslexia were ignored. In our view, this is troubling, since these are very important guidelines and they could contribute significantly to easier learning of people with dyslexia from the materials.

Table 2
APPLYING ACCESSIBILITY GUIDELINES

Accessibility Indicators	%
2. The material contains navigation	100.0
3. The navigation is clearly separated from the content	100.0
7. The material does not consist of text only	100.0
9. The content is supported with pictures, charts, illustrations	98.0
13. Pictures, illustrations, diagrams have their equivalents	83.7
19. Audio and video recordings and animations can be replayed by the user	79.6
18. Audio and video recordings and animations can be paused by the user.)	73.5
17. Audio and video recordings and animations can be started by the user	69.4
20. The material does not contain flashing elements that cannot be stopped	65.3
10. The content is supported with animations	63.3
16. Animations have their equivalents	51.0
12. The content is supported with video recordings	40.8
4. The material contains controllers to regulate font sizes	34.7
11. The content is supported with audio recordings	30.6
15. Video recordings have their equivalents	22.4
14. Audio recordings have their equivalents	18.4
1. The material enables access via assistive technology (text to speech)	0.0
5. The material contains controllers to regulate typefaces	0.0
6. The material contains controllers to regulate background and text colours	0.0
8. The material enables listening to the whole of material	0.0

Text is often a source of anxiety for people with dyslexia, and if it is presented in unsuitable or distracting size, typeface or colour, sharp contrasts between background and typeface colours, it makes it even more frustrating for them. Today's technology allows installing controllers, which enable the customization of typefaces and font sizes as well as background/text colours. We established that text customisation is usually limited to font size, whereas typefaces and background/text colours cannot be set by the user. The possibility of modifying text appearance may be crucial to people with dyslexia and their learning. Thus, it is necessary that they offer such adjustments in web-based learning materials. Materials with audio options are a great help to many a person with dyslexia. Not even one of the examined web-based learning materials enabled, in its entirety, to be "listened to". However, some web-based learning materials did permit listening to parts of their contents, which is certainly positive.

While images typically had their textual equivalents, we noticed that this was only rarely true of video and audio recordings. Relying on only visuals is more of a disadvantage than an advantage for people with dyslexia. In addition to visual/audio presentation, it is always better to add written explanation to support what is seen/heard. Ensuring equivalent descriptions is also important when considering access to the content via assistive technology (text-to-speech converters). The latter cannot read video or audio recordings, so the user who does not read the text cannot get the same information from such web-based learning materials as someone who reads the text.

5.2.2 Usability

We will now look into how the guidelines were applied in the area of usability.

Table 3
APPLYING USABILITY GUIDELINES

Usability Indicators	%
1. The material uses a sans serif typeface (Verdana, Arial, Georgia etc.)	100.0
4. The site map is hierarchical and it gives an overview of the complete material.	100.0
7. The material contains a progress indicator	100.0
8. When we want to return to the beginning or to a specific section, we do not have to go through all the sections.	100.0
9. The form and navigation of the material remain consistent throughout the material.	100.0
13. The content is provided on one screen with minimal vertical scrolling.	100.0
15. The material is divided into short paragraphs.	100.0
16. The text is in the middle or on the right side of the screen. It does not take up the whole screen.	98.0
14. There is no horizontal scrolling.	91.8
12. Hyperlinks are descriptive and we know where they will take us.	87.8
5. The material contains navigation forward/back buttons.	85.7

Usability Indicators	%
17. The material uses left text justification.	85.7
2. The font size is at least 12 pt.	59.2
18. The text is not crammed; it uses adequate line spacing (at least 1.5).	55.1
10. Textual hyperlinks are coloured when they have been clicked.	0.0

As we noticed when addressing the first research question, usability is the category that the developers of web-based learning materials take into account the most. The only serious problem we detected was the guideline related to hyperlinks, namely that textual hyperlinks should be coloured when they have been clicked. It means that no material with textual hyperlinks allows them to be coloured when they have been clicked. Since people with dyslexia may have memory problems and become lost quickly within web-based learning materials, consequently forgetting which link they have already clicked, a good and quick solution is allowing for the links that have been clicked to be coloured. This is the way for people with dyslexia to know which hyperlinks they have already visited. In addition, if returning to the same web-based learning materials they will have a good overview of what contents they have already studied.

5.2.3 Readability

Finally, we would like to consider how readability guidelines were applied.

Table 4
APPLYING READABILITY GUIDELINES

Readability Indicators	%
4. The text has clearly visible headings and subheadings.	100,0
8. Graphics are used better to illustrate and explain any complex text.	100,0
13. The material does not contain large chunks of underlined text, which is a not hyperlink.	100,0
12. The material uses the active voice, not the passive.	98,0
7. Bullet points are used for better clarity.	95,9
1. The average sentence length is between 15 and 20 words.	93,9
5. Important information is in bold or highlighted.	93,9
3. Instructions are given clearly and without lengthy explanations.	87,8
2. At the beginning, key information is emphasized (learning objectives, expectations...)	73,5
11. The material provides suggestions, additional explanations and links at the side or in drop-down menus or when moving the mouse over a text.	34.7
9. New concepts are explained in glossaries, icons in legends.	26.5
10. Boxes and mind maps are used to summarize important points.	12.2

Despite the relatively good result in this area, there remain three guidelines in the readability category with low scores. As the table shows, the smallest percentage was attained by the guideline *boxes and mind maps are used to summarise important points* (12.2%). As many people with dyslexia have problems with structure, summaries are a very welcome solution, making it easier for them to orient themselves and to revise what they have already learnt (Reid, 2009). Mind maps and boxes as a form of presenting summaries are especially emphasised, because visual support is particularly welcomed by such individuals.

Readability is also affected by the understanding of text. This is related to other two guidelines (11 and 9), that are both important to e-learning: if words or explanations are not understood in traditional learning in the classroom, we can ask the teacher to explain them again, but we cannot do so when learning from web-based learning materials. Therefore, it is important that web-based learning materials provide certain explanations that will function in a similar way as the teacher in the classroom. Because of reversing and inverting letters, people with dyslexia sometimes read words wrongly (especially more complex, longer or foreign words, etc.), so it is right to facilitate their reading by providing additional explanation in the form of a glossary or within the text itself (explanation at the side or when moving the mouse over a text, etc.). In web-based learning materials, such adaptations are extremely simple and should be applied as much as possible.

Conclusion

Although multimedia has become an important part of our knowledge-based society, the abilities to read and write remain crucial to understanding complex materials. They, furthermore, are preconditions for social and, especially, “digital” integration (Torrisi & Piangerelli, 2010). Due to the struggles, they have to face when facing web-based learning materials, people with dyslexia remain a marginalised group. Although reading and writing difficulties can be compensated to a degree with the help of various technologies, the problems can persist if specific adaptations are not provided. Thus, it is vital that the developers of online services and, particularly, materials for online learning realise what problems this group of individuals’ faces, and avoid creating the so-called “exclusive digital environments” (Monteiro & Leite, 2016).

In its essence e-learning strives to ensure the suitability of the learning process for the individual’s needs, goals and wishes and to enable access to knowledge at the time, at the place and in the manner suitable for her/his needs. Although technological and learning support is undoubtedly one of the important factors in the success and effectiveness of any learning (Radovan

& Makovec, 2015), it is critical that we look for reasons deeper than that, too. Romiszowski (2000) lists as one of the most important reasons for failures in e-learning, also a failure to take account of participants' needs. According to the findings of our research study, we can conclude that a somewhat poorer application of certain guidelines important for dyslexic users may also be blamed on the lack of awareness among the developers of the needs of such users.

Such a result is also a consequence of the fact that dyslexia as a specific learning disability is only rarely addressed independently; rather, it is approached within the group of cognitive problems, without being dealt with separately. Literature and research on accessibility, also, primarily focus on physically disabled, blind and deaf individuals (Freire, Petrie, & Power, 2011), but dyslexia does not seem to figure appropriately in these contexts.

It often turns out that e-learning material developers are technologists and designers, who want to make the materials as visually attractive as possible, often neglecting these criteria. Moreover, the general accessibility criteria and guidelines developed by the Web Accessibility Initiative (WAI) are written in such a complex and incomprehensible way that even designers themselves find them hard to understand, let alone the teachers who set about compiling e-learning contents. Consequently, we should pay special attention to drawing up clear accessibility guidelines in order to make them readily comprehensible to everybody. More engagement from policymakers would certainly signify a step forward, promoting as they should a legally enforced right to these adaptations in e-learning. This would eventually have a positive influence not only on people with dyslexia but also on those who do not have dyslexia.

REFERENCES

Agrusti, G. (2013), Inquiry-based learning in Science Education. Why e-learning can make a difference. *Journal of E-Learning and Knowledge Society, 9*(2). https://doi.org/10.20368/1971-8829/831

Burt, P. (2004), *Design guidelines for effective e-learning materials*. Retrieved from http://www.paulburt.co.uk/e-learning/e-learn_design.pdf

Camp, D., & Aldridge, J. (2007), Rethinking dyslexia, scripted reading, and federal mandates: The more things change, the more they stay the same. Journal of Instructional Psychology, 34(1), 3–12.

Dinevski, D., & Radovan, M. (2013), Adult Learning and the Promise of New Technologies. *New Directions for Adult and Continuing Education*, (138), 61–69. https://doi.org/10.1002/ace.20054

Doyle, J. (2002), *Dyslexia: An introductory guide*. London: Whurr Publishers.

Dunn, S. (2003), *Return to SEDNA? Implementing accessibility for disabled students* *in virtual learning environments in UK further and higher education*. London: City University. Retrieved from http://www.saradunn.net/VLEproject

Fee, K. (2009), *Delivering e-learning. A complete strategy for design, application, and assessment*. Philadelphia: Kogan Page.

Freire, A. P., Petrie, H., & Power, C. D. (2011), *Empirical results from an evaluation of the accessibility of websites by dyslexic users* [Proceedings Paper]. Retrieved 27 June 2017, from http://eprints.whiterose.ac.uk/109244/

Mažgon, J., Šebart, M. K., & Štefanc, D. (2015), The Role and Use of E-Materials in Vocational Education and Training: The Case of Slovenia. *Turkish Online Journal of Educational Technology - TOJET, 14*(4), 157–164.

McCarthy, J. E., & Swierenga, S. J. (2010), What we know about dyslexia and Web accessibility: a research review. *Universal Access in the Information Society*, *9*(2), 147–152. Retrieved from http://doi.org/10.1007/s10209-009-0160-5

Monteiro, A., & Leite, C. (2016). Inclusive digital online environments as a device for pedagogic differentiation: a taxonomy proposal. *Journal of E-Learning and Knowledge Society*, *12*(4). https://doi.org/10.20368/1971-8829/1115

Radovan, M., & Makovec, D. (2015), Relations between Students' Motivation, and Perceptions of the Learning Environment. *Center for Educational Policy Studies Journal, 5*(2), 115–138.

Radovan, M., & Perdih, M. (2016), Developing Guidelines for Evaluating the Adaptation of Accessible Web-Based Learning Materials. *The International Review of Research in Open and Distributed Learning*, *17*(4). https://doi.org/10.19173/irrodl.v17i4.2463

Rainger, P. (2003), A dyslexic perspective on e-content accessibility. *TechDis*. Retrieved from http://www.techdis.ac.uk/seven/papers

Reid, G. (2009), *Dyslexia: A practitioner's handbook* (4th ed). Malden, Mass: Wiley-Blackwell.

Smythe, I. (2010), *Dyslexia in the digital age: Making IT work*. New York: Continuum International Publishing Group.

Smythe, I., & Draffan, E. A. (2005), Dyslexia and e-learning – a guide to good practice. In I. Smythe (Ed.), *Provision and use of Information technology with dyslexic students in university in Europe* (pp. 79–86). Retrieved from http://eprints.soton.ac.uk/264150/1/The_Book.pdf

Torrisi, G., & Piangerelli, S. (2010), How new technologies can help with "invisible disabilities." *eLearning Papers*, (19), 1–7.