

# Web Service to Automate Bibliographic Research

## Case of Dependability Ontology

<https://doi.org/10.3991/ijes.v8i2.15067>

Fatima Ezzahra Mdarbi <sup>(✉)</sup>, Nadia Zahir  
Hassan II University of Casablanca, Casablanca, Morocco  
fatimzahra.mdarbi@ensem.ac.ma

Nadia Afifi, Imane Hilal  
Hassan II University of Casablanca, Casablanca, Morocco  
School of Information Sciences, Rabat, Morocco

**Abstract**—In order to identify the existence of an ontology of dependability, we conducted a bibliographic review. However, we encountered a real problem while researching and selecting articles that address this issue. Indeed the research process in various scientific databases with standard techniques involve the application of a number of manual and duplicative steps. This makes it a fairly costly and faulty process. To surmount these challenges, we offer a solution that automates search and selection items. In this article, we propose a web service that automates the process of searching in scientific databases in order to extract relevant papers about dependability's ontology. The proposed solution is essentially derived from the popular method called systematic mapping. This Web Service will allow users to query scientific databases to obtain the metadata of selected articles. Our proposal will make the selection of scientific articles easier and faster.

**Keywords**—Ontology, dependability, web service, bibliographic research, systematic mapping

## 1 Introduction

Nowadays, resource misuse and system vulnerability are often observed, due to

1. Increasing data volume
2. Resources heterogeneity
3. Dynamic nature of massive user requests

Achieving an adequate degree of certainty for systems remains a major concern. Ontology is characterized as a model of a common conception. It provides semantics through relations between concepts and can be applied widely to data from heterogeneous resources.

Ontologies are the most widely used representation in engineering. They intend to set up a representation where machines can manipulate the semiology of information ontologies, terminologies, classifications and coding systems can be identified as terminology resources. Ontologies and other terminology resources are become more significant in semantic alignment for data integration, information exchange and semantic interoperability. For Obitko et al. ontologies are essential for representing and exchanging knowledge. In this context, they present a method for identifying conceptual structures among data sets. This solution uncovers the need for new contexts and relations in an ontology [1].

Our main objective is to identify the availability of a specified dependability ontology. But with the different concepts of dependability, a simple bibliography was not enough to prove it. In this context we conduct a literature search for scientific articles to have an overview in this area.

During our research for existing dependability ontology, we notice that the research with standard techniques involve the application of a number of manual and duplicative steps. To overcome this problem, we have proposed an automatic process that makes it easier to search for and find articles using Systematic Mapping (SM).

This article is structured such as: we present in Section 2 related work to scientific search methods. In section 3, we present the terminology related to our implementation. The main expected functionalities for our solution are describe in Section 4, while implementation details will be discussed in Section 5. Finally, in Section 6, we synthesize with a summary and perspectives for our future work.

## 2 Related Work to Scientific Search Methods

Numerous scientific articles address the dependability ontology. Therefore, identifying and investigating all primary studies manually requires a series of iterative steps. To address these challenges, we searched for an efficient existing research method.

Several guidelines have been developed to encourage the use of secondary studies [2]. Existing guidelines include: string research in databases [3], and Snowballing [4], [5]. There are other forms of search, presented as supplementary to the suggested guidelines.

Some guidelines introduced the text extraction as an interesting supplementary selection procedure for searching [6]. An automated approach to search and select was suggested on this basis [7].

The comparison of the use of text for researching in databases and snowballing in software developing rea [8], has demonstrated that the two approaches are comparably, and as both made a reasonably sample of pertinent surveys. In addition, the snowballing method was employed by some researchers as an initial research method and then supplemented by a database search to assess the utility of the snowballing method [9]. In addition, they confirm that the initial method may be potentially much secure; this depends heavily on the construction of a suitable starter set.

Same researchers have presented an approach to automate research based on snowballing [10].The snowballing reliability is more essential in the context of second-

generation systematic reviews [5]. This procedure proved to be the mostly adequate since current studies will refer to earlier ones.

The univariate research techniques are the most simple optimization algorithm to deploy.

When all parameters have been modified, the entire process is replicated until the solution objective is accomplished or a specified iteration number is achieved

The parameters variation number indicates the search intensity. While the loops in all parameters number represents the search scope.

A weighting method often used in information retrieval and particularly in text mining known as term frequency-inverse document frequency (TF-IDF). It is used to assess the relevance of a word or expression in a document in relation to a particular collection. The weights are increased in proportionality to the number of times the word occurs in the document [11].

An examination in [12] of results of applying TF-IDF, in order to determine what words in a corpus of documents might be more favorable to use in a query. This examination provide evidence that this simple algorithm efficiently categorizes relevant words that can enhance query retrieval.

However, the traditional TF-IDF is not fully effective for text classification. In this context authors in [13] carry out a comparable study on the various term weighting systems and suggest a new one. Term frequency & inverse gravity moment (TF-IGM) TF-IGM integrates a special model to accurately quantify the term distinction class. [14] Using Floating search methods in feature selection.

The objective is to choose a subset of  $d$  characteristics from the defined set of measurements  $D$ ,  $d < D$ , while not significantly affecting the system efficiency.

Direct search methods (DSM) are a category of optimization procedures that do not explicitly use derivative. Rather, they operate directly with the objective function to conduct the search for an optimized value. In recent years, DSM have attracted increased interest because of new mathematical analysis, their relevance to both parallel and distributed computation, and their usefulness in solving optimization challenges that generate complicated computing solutions [15].

Authors in [16] discussing the DSM for unrestricted optimization. They provide a current insight into this classical algorithm and examine how direct search methods are distinguished by the lack of constructing an objective model.

In our case, to identify the existence or not of an ontology of dependability, we had to apply the research methods mentioned above. However, despite the theoretical efficiency of all these methods, we had to perform a series of repetitive and tedious manual steps for each, as none of them offers the possibility of automating the bibliographical research process.

### 3 Terminologies: Dependability, Ontology, and Research Methods

#### 3.1 Dependability

Dependability is about assessing potential risks, anticipating failure and minimizing the effects of a disaster when it is encountered [17].

Dependability is the system ability to enable users to build an appropriate degree of confidence in the service provided [18].

Dependability is the capacity of entities to achieve several functions under specific requirements. It can be performed at the process, system or component level, according to the level of analysis. The concepts to operational safety are grouped into three fundamental concepts as seen in figure 1.

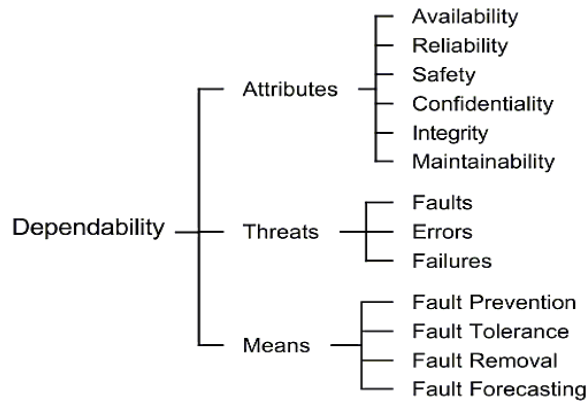


Fig. 1. Dependability tree

In our previous works [19], we have encountered several obstacles when dealing with the concepts of dependability. The variety of dependability's concepts, require establishing a coherent semantic representation using a strong modelling. Ontologies are among the most used models for field representation. Therefore, having an ontology of dependability constitutes an important step to formalize the relationship between its concepts.

#### 3.2 Ontology

An ontology is a structured and explicit formulation of a common conception. An ontology is a conceptual formulation describing the domain concept independently [11]. An ontology has two possible ways of constructing: domain-dependent and generic [12]. The ontological approach allows a common and shared understanding between all the actors of the Semantic web [13].

The ontologies used for describing and structuring resources will be presented in a various formats and with different languages [14]. Thus, ontology is used in different

areas and for different purposes [15], and is described as a set of primary terminology used in creating artificial systems [16].

As ontology is a specific form of knowledge model [12], this knowledge generated by an ontology must be diffused following five elements [8]:

- Concepts: represent pertinent abstractions of reality, selected based on specific purpose and desired application of the ontology.
- Relations: represent the current combinations of the concepts in the reality segment. These relations permit us seeing the structuration and interrelation of the concepts between them.
- Functions: represent a special case of relationships. Where the n-th element of a relationship is determined based on the n-1 preceding elements.
- Axioms: represent the inference laws for specifying some characteristics of relationships.
- Instances: represent the extensional definition of ontology.

To benefit from the strengths of ontology in the case of dependability, we must first make sure that it exists or not. To do this, we have conducted a literature review of articles that deal with this topic. During this study, we used existing scientific research methods to have an overview of this area.

### 3.3 Scientific search methods

In order to get an idea of research field, researchers employ secondary education. A secondary study reviews all primary studies linked to a selected research question in order to obtain integrity and summarize the evidence on that question

Two different methods can be distinguished from existing secondary studies: (i) systematic review (SR) and (ii) systematic mapping (SM). These methods follow a strictly defined procedure reducing the risk in their conclusions [20].

**Systematic review:** A systematic literature review is a method of identifying, evaluating and interpreting existing literature related to a specific search, thematic area or interest issue. In addition, surveys that contribute to a systematic review are known as primary studies; a systematic review is a version of a secondary study.

**Systematic mapping:** SM is conceived to offer a survey of a research topic. The SM results may identify fields that are appropriate for performing SR and areas where a primary study is more suitable.

**Systematic mapping vs systematic review:** The main differences between a SM and SR are:

- SM is usually guided by research questioning and often involve a number of research questions.
- Keywords for SM will be less targeted than SR and are more susceptible to return a very high number of studies.
- The procedure of extracting data for the SM is larger than that of the RS and can be more specifically described as a classification or categorization step. The purpose of this step is to categorize the articles in adequate detail.

## **4 Proposed Solution**

As for our study, we adopt the SM process; we propose an automatic process to search in scientific databases in order to extract relevant papers about dependability's ontology.

The initial idea is based on the classic approach proposed by Kitchenham and Charters [21]. The keyword identification followed the research questions. For every source, we suggest a triple search for the keywords you are looking for: in the title, in the abstract and in the keywords. Then, we applied the article evaluation criteria to classify the most pertinent articles we apply the paper selection criteria to categorize the most relevant papers. A ranking is made for every class by subject area. We determine the category that correspond to our area.

We compared the keywords determined for the search and the keywords in the article and then calculated the number of similarities between them. The selected paper is more significant if the correspondence is higher.

Our proposed solution uses R packages to extract the meta-data's articles from different paper's sources. Then, the solution employs an abstracted interface, which supports Web Service technologies.

### **4.1 Sources of papers**

Previously, the academic community had to search manually for articles by visiting libraries. In the early 1990s, access to the full text of articles, including conference proceedings, became possible through the Internet. The availability of search engines such us: Google, Bing, Yahoo, made it possible for everyone to search in the Internet rapidly and practically [22].

As students' progress in their studies, they are subjected to more advanced topics. The use of scientific databases becomes essential to further investigate the topics being researched [23].

Despite the presence of many credible and indexed, scholarly databases and source engines, there are many access restrictions due to author's copyrights and commercial strategies.

Among our proposed approach entries, we have papers sources including scholarly databases and source engines. Some scientific databases expose many of their metadata's articles under API. Although, most of this APIs requires API key and users have to show their intention behind using the API. To interact with this APIs, scientists use the R language.

### **4.2 R language**

The R language was developed by Ross Ihaka and Robert Gentleman at the University of Auckland in the 1990, and it is a derivative of the S language [24], [25].

The language has grown in popularity over the years [36], with applications that include data discovering and analysis.

To build our method implementation, we aim to use R packages to extract meta-data from databases and present a flexible solution to final users by using web technologies.

### 4.3 Web services technology

Web technologies are the standards techniques for Internet application. It uses simple and relatively open sets of standards such: HTTP, HTML, XML, JSON and TCP/IP. The principal goal is to facilitate the construction of scalable and coupled services.

The web technologies utilization minimize the difficulty of developing and implementing a distributed platform. This makes it more usable for different systems, such as desktops, laptops, smart phones, tablets and video game consoles.

A web service can be defined as a software entity that can provide features on network. One of the most widely used web service applications is the SOAP-based web service. This conception comprises (i) Universal Description Discovery and Integration (UDDI), (ii) Web Services Description Language (WSDL), and (iii) SOAP protocols.

- UDDI is an extendable markup language protocol that contains a process for storing, indexing, and localizing web service applications.
- WSDL is an XML-based interface definition language used for describing the functionality offered by a Web Service.
- SOAP is a protocol that specified the message format for exchanging structured information.

SOAP and WSDL provide basic message exchange. The increasing use of Web Services made the necessity for additional functionalities such as: security management, transaction management, message reliability. This need has conducted to the WS-\*emergence.

WS-\* are the second-generation Web Services specifications. They extend the specifications of the first generation of specifications consisting of SOAP, WSDL, and UDDI.

Representational State Transfer (REST) is an architecture design developed by Roy Fielding for distributing systems such as the Internet [26]. This design defines six restrictions: (i) Client-server; (ii) Stateless; (iii) Cacheable; (iv) Layered system; (v) Code on demand (optional) ; and (vi) Uniform interface. Information and representations are the main features of RESTful Web Services. Information and representations are the main features of RESTful Web Services.

If browser submits an HTTP GET request to a web application for specific information, it obtains an HTML response as a resources representation. The HTTP methods, for CRUD (Create, Read, Update and Delete), used by RESTful Web Services are: (i)GET, (ii) PUT, (iii)POST, and (iv) DELETE.

The RESTful design contain the following four features:

1. One URI is related to one resource.
2. Standard methods are applied: GET, POST, PUT, and DELETE.
3. Various resource states are sent to the customer.
4. Communications are static.

Table 1 presents the differences between the two architectural styles WS-\* and REST [27].

**Table 1.** Comparative between REST and WS-\*

Property	REST	WS-*
Adress to Ressource/Service	One-to-One	One-to-many
General Operations	HTTP (GET,POST,PUT,DELETE)	Self-Defined
Communiacion Protocols	HTTP	HTTP/SMTP
Data Formats	TEXT,JSON,XML	XML
Communication Context	None	XML Schema ,WSDL UDDI
Complexity	Low	High
Cohesion Degree	Loose	Tight
Application Scale	Large	Medium,small
Scenarios	Lighrweight,Compatible, Unified	Transactional,Secure, Coordinated

The open and expandable architecture normally requires features as: i) the system be less complex, (ii) easy access, (iii) lightweight, (iv) scalable, and (v) flexible.

(i)

Those features are present in the RESTful design. Due to its advantages, RESTful design remains an interesting web technology used in the R environment and system; such as the Opencpu project, which aims to facilitate the use of R within a designing interface, especially for native R users.

#### 4.4 The OpenCPU system

The OpenCPU software is presented as an implementation that builds on HTTP and R to expose an abstracted interface. OpenCPU defines an HTTP API developed on R language. The API is designed to describe general logic of data analysis.

Two OpenCPU server implementations were developed: (i) The first one is the R package OpenCPU which uses the httpuv Web Server. (ii)The second one is the cloud server which is a multi-user implementation based on Ubuntu Linux and rApache. This cloud server provides better performance, advanced security and more configuration options, however it requires a dedicated Linux server.

In addition, the main difference between these implementations is the way they manage competition. Since R is a unique thread, httpuv only processes one query at a time. Additional requests are automatically queued and executed in succession using the same process.

The API provides a standard interface to the both implementations(R package OpenCPU and OpenCPU cloud server). Depending on the needed performance, user's applications will behave the same regardless of which server is used. All the previous technologies are presented in order to build the components of our §solution



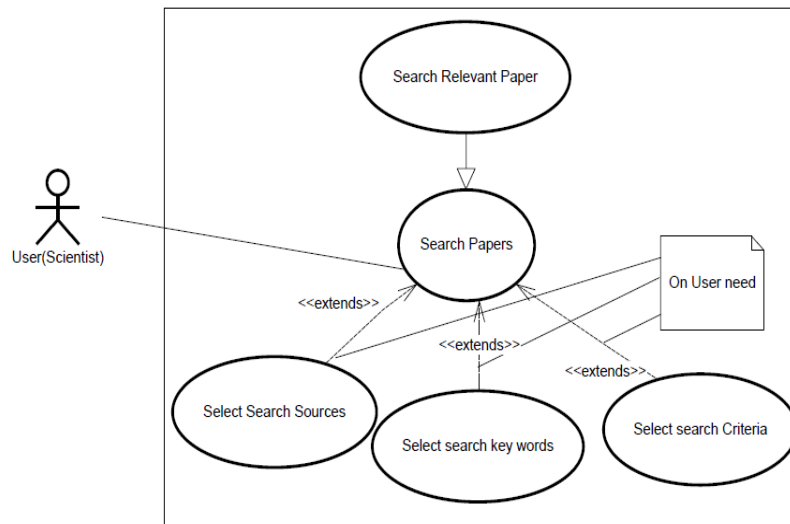
## 5 Solution Implementation

### 5.1 Solution modeling

In order to deliver a multiple source interface to scientific papers. We implement a solution based on a Web Services that uses R packages, and interact with APIs provided by scientific databases.

The main purpose is to implement our proposed search method [28], in which user start by choosing a topic, selecting key words and selecting research sources, the system will execute the process and output relevant paper according to user entries.

Users also can get articles corresponding to their custom criteria. The Use Cases for the expected functionality are illustrated in figure 2.



**Fig. 2.** User use cases

In order to return relevant papers, many exchanges are preceded in our system as shown in figure 3:

- The user select his entries (Key words, data Sources, Criteria) using his browser
- The Browse send an Ajax call to the server via a Web Service
- The Web Service make remote calls to R packages corresponding to user selected sources
- Each R package interact with Scientific Database API to get xml metadata
- The R package returns its R response and sends it to Web Service.
- The Web Service adapts package's response and sends it to browser in JSON format.
- Finally, the browser shows results to the user.

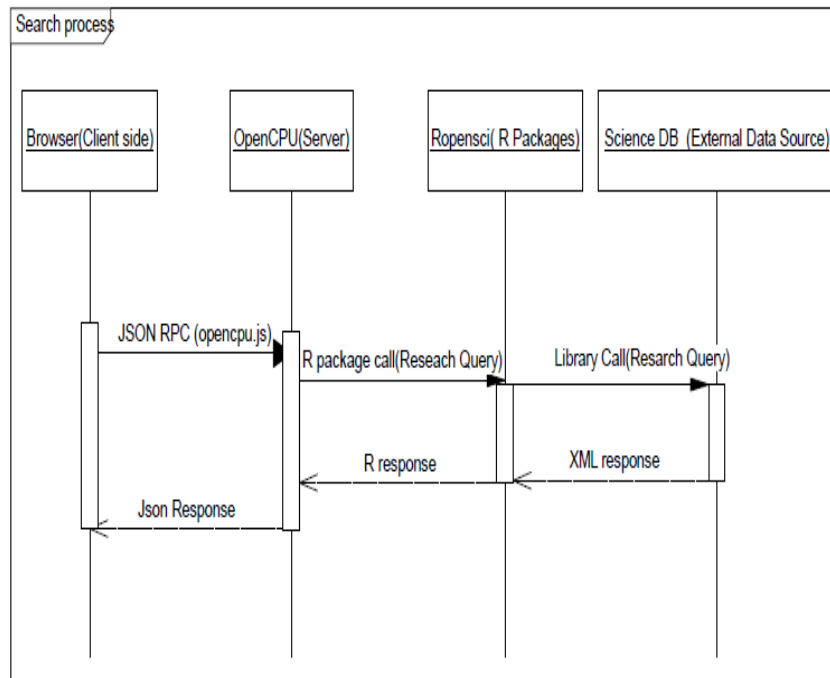


Fig. 3. Sequence diagramme for search process on the system

## 5.2 Implementation

The user connects to the Web server OpenCPU and invokes the service dedicated. After analyzing the query sent by the user, the Web Service calls the search R package responsible for translating the user query to different R packages. The Web Service performs remote calls to the set of R packages, and each R package connects to scientific database's API to get full text or metadata of available articles. As shown in figure 4.

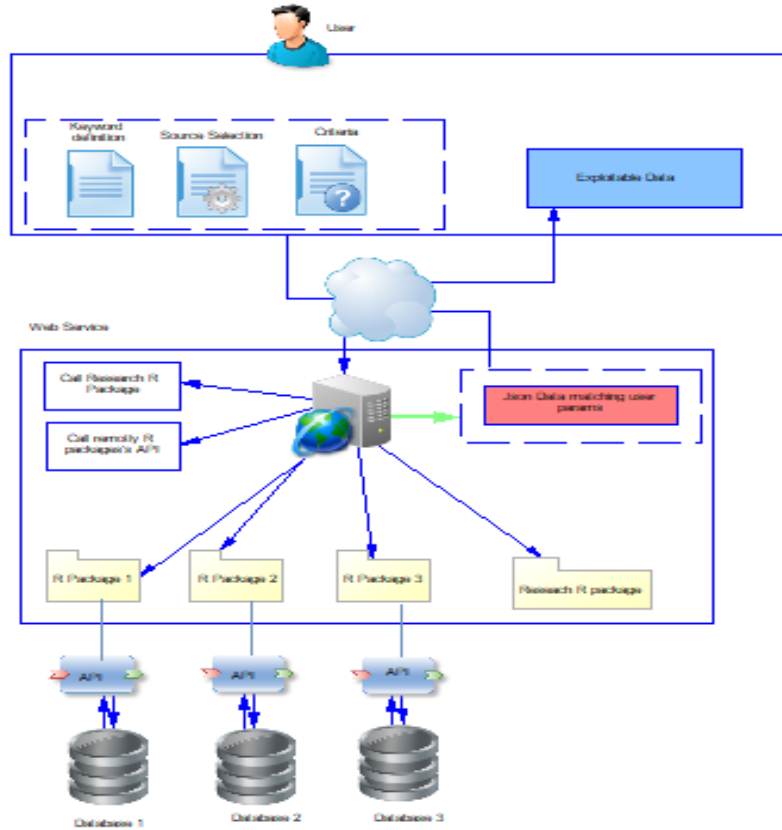


Fig. 4. Process invoking a Web service to get metadata’s article

Moreover, we can get full text article or article meta-data using specific R methods and store results in user disc.

For the client side, in order to invoke the R research package from browser we use the `ocpu.rcp` method from the `opencpu.js` of `Opencpu`; it is a library builds on `jQuery` to provide Ajax wrappers for calling R within a web page. So finally, user will have on output, articles meta-data and can explore full text article according to his need.

To identify all works in a subject, we adopt a search by its keyword in multiple sources. We represent the set of dois returned from the source S as (1):

$$S_{dois} = \{RSk1 \cup RSk2 \cup RSk3 \cup \dots \cup RSk_n\} \tag{1}$$

Where  $RSk_i$  refers to Result Set of DOI returned by the Source S for the given keyword  $k_i$ .

Therefore, to have the integrity of articles from different sources we perform the union of all result set from n sources as (2):

$$R_{dois} = \cup S_{dois} \tag{2}$$

The example below figure 5 is an example of implementation of our solution. It shows an R code from the R Research package, it includes an R function for searching in Plos scientific library. It uses the “Fulltext”, an R package that interact with Plos API. The example addresses the case of two keywords: the first one is dependability, and the second is ontology.

The request aims to get articles identifiers for each keyword from the specified source before returning the intersection of found articles identifiers. Thus, the search function returns articles that have both ontology and dependability in their full text or meta-data depending on the source API.

```
keyWordFirst <- "Dependability"
keyWordSecond <- "Ontology"
plosSearch <- function (keyWordFirst, keyWordSecond,
limitResult)
{
  #Search Query
  resultatKeyWordFirst<-ft_search(query=keyWordFirst,
from='plos',
  limit=limitResult,
  plosopts=list (start=0, limit=limitResult));
  #List of found articles doi
  doisKeyWordFirst<-resultatKeyWordFirst$plos$data$id;
  #Search Query
  resultatkeyWordSecond<-ft_search (query=keyWordSecond,
from='plos', limit=limitResult,
  plosopts=list (start=0, limit=1000)) ;
  #List of found articles doi
  doiskeyWorddaSecond<-resultatkeyWordSec-
ond$plos$data$id ;
  #Intersection of found dois
  intersectKeysWords <- intersect (doisKeyWordFirst,doi-
skeyWordSecond);
  # Result return
  return (intersectKeysWords)
}
```

Fig. 5. Listing code

## 6 Conclusion and Perspectives

Standard research methods require fastidious and painful steps for scientific researchers. Therefore, a solution that supported all or part of research mechanism is crucial as alternative. In this article, we propose a Web Service that automate research bibliographic. This Web Service is based on SM method, and use R language and the string search method in scientific databases. It will ensure pertinent articles and

coherent research. To prove its efficiency, we apply it to the case study dealing with dependability ontology of. Finally, we get a pertinent and a consistent result.

As perspective, our proposal will simplify the step of exploiting the previous works on ontologies of dependability in order to construct a new one adaptable later for Big Data area ref cloud and join our previous works. We intend to join our solution to our previous work dealing with cloud service selection for Big Data[29].

## 7 References

- [1] M. Obitko, V. Snasel, J. Smid, et V. Snasel, « Ontology Design with Formal Concept Analysis. », in CLA, 2004, vol. 128, no 3, p. 1377–90.
- [2] V. Nepomuceno et S. Soares, « On the need to update systematic literature reviews », Information and Software Technology, vol. 109, p. 40-42, mai 2019, <https://doi.org/10.1016/j.infsof.2019.01.005>.
- [3] J. Brings, M. Daun, M. Kempe, et T. Weyer, « On Different Search Methods for Systematic Literature Reviews and Maps: Experiences from a Literature Search on Validation and Verification of Emergent Behavior », in Proceedings of the 22Nd International Conference on Evaluation and Assessment in Software Engineering 2018, New York, NY, USA, 2018, p. 35–45, <https://doi.org/10.1145/3210459.3210463>.
- [4] K. R. Felizardo, A. Y. I. da Silva, É. F. de Souza, N. L. Vijaykumar, et E. Y. Nakagawa, « Evaluating Strategies for Forward Snowballing Application to Support Secondary Studies Updates: Emergent Results », in Proceedings of the XXXII Brazilian Symposium on Software Engineering, New York, NY, USA, 2018, p. 184–189, <https://doi.org/10.1145/3266237.3266240>.
- [5] C. Wohlin, « Second-generation systematic literature studies using snowballing », in Proceedings of the 20th International Conference on Evaluation and Assessment in Software Engineering, 2016, p. 15, Consulté le: janv. 07, 2017. [En ligne]. Disponible sur: <http://dl.acm.org/citation.cfm?id=2916006>. <https://doi.org/10.1145/2915970.2916006>
- [6] K. R. Felizardo, S. R. S. Souza, et J. C. Maldonado, « The Use of Visual Text Mining to Support the Study Selection Activity in Systematic Literature Reviews: A Replication Study », in 2013 3rd International Workshop on Replication in Empirical Software Engineering Research, oct. 2013, p. 91-100, <https://doi.org/10.1109/reser.2013.9>
- [7] F. Tomassetti, G. Rizzo, A. Vetro, L. Ardito, M. Torchiano, et M. Morisio, « Linked data approach for selection process automation in systematic reviews », in 15th Annual Conference on Evaluation & Assessment in Software Engineering (EASE 2011), 2011, p. 31–35. <https://doi.org/10.1049/ic.2011.0004>
- [8] S. Jalali et C. Wohlin, « Systematic literature studies: Database searches vs. backward snowballing », in Proceedings of the 2012 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement, sept. 2012, p. 29 - 38, <https://doi.org/10.1145/2372251.2372257>.
- [9] D. Badampudi, C. Wohlin, et K. Petersen, « Experiences from Using Snowballing and Database Searches in Systematic Literature Studies », in Proceedings of the 19th International Conference on Evaluation and Assessment in Software Engineering, New York, NY, USA, 2015, p. 17:1–17:10, <https://doi.org/10.1145/2745802.2745818>.
- [10] F. Bezerra, C. H. Favacho, R. Souza, et C. de Souza, « Towards Supporting Systematic Mappings Studies: An Automatic Snowballing Approach », p. 11, 2014.

- [11] « TF-IDF », Wikipédia. févr. 10, 2019, Consulté le: janv. 08, 2020. [En ligne]. Disponible sur: <https://fr.wikipedia.org/w/index.php?title=TF-IDF&oldid=156618833>.
- [12] « Using TF-IDF to determine word relevance in document queries », ResearchGate. [https://www.researchgate.net/publication/228818851\\_Using\\_TF-IDF\\_to\\_determine\\_word\\_relevance\\_in\\_document\\_queries](https://www.researchgate.net/publication/228818851_Using_TF-IDF_to_determine_word_relevance_in_document_queries) (consulté le janv. 08, 2020). <https://doi.org/10.18535/ijecs/v5i4.40>
- [13] K. Chen, Z. Zhang, J. Long, et H. Zhang, « Turning from TF-IDF to TF-IGM for term weighting in text classification », *Expert Systems with Applications*, vol. 66, p. 245-260, déc. 2016, <https://doi.org/10.1016/j.eswa.2016.09.009>.
- [14] P. Pudil, J. Novovičová, et J. Kittler, « Floating search methods in feature selection », *Pattern Recognition Letters*, vol. 15, no 11, p. 1119-1125, nov. 1994, [https://doi.org/10.1016/0167-8655\(94\)90127-9](https://doi.org/10.1016/0167-8655(94)90127-9).
- [15] R. M. Lewis et V. Torczon, « Direct Search Methods », in *Wiley Encyclopedia of Operations Research and Management Science*, American Cancer Society, 2011.
- [16] R. M. Lewis, V. Torczon, et M. W. Trosset, « Direct search methods: then and now », *Journal of Computational and Applied Mathematics*, vol. 124, no 1, p. 191-207, déc. 2000, [https://doi.org/10.1016/s0377-0427\(00\)00423-4](https://doi.org/10.1016/s0377-0427(00)00423-4).
- [17] M. Fatima Ezzahra, A. Nadia, et H. Imane, « Big Data Dependability Opportunities Challenges », in *2019 1st International Conference on Smart Systems and Data Science (ICSSD)*, oct. 2019, p. 1-4, <https://doi.org/10.1109/icssd47982.2019.9002676>.
- [18] A. Aviz, J.-C. Laprie, et B. Randell, « Fundamental Concepts of Dependability », p. 21.
- [19] I. Hilal, N. Affi, R. Filali Hilali, et M. Ouzif, « Toward a New Approach for Modeling Dependability of Data Warehouse System », *International Journal of Computer Science and Information Security*, vol. 11, no 6, p. 31051344, juin 2013.
- [20] K. R. F. Scannavino, « Evidence-based software engineering: systematic literature review process based on visual text mining », text, Universidade de São Paulo, 2012. <https://doi.org/10.11606/t.55.2012.tde-18072012-102032>
- [21] S. Keele, « Guidelines for performing systematic literature reviews in software engineering », Technical report, Ver. 2.3 EBSE Technical Report. EBSE, 2007.
- [22] S. Büttcher, C. L. Clarke, et G. V. Cormack, *Information retrieval: Implementing and evaluating search engines*. Mit Press, 2016.
- [23] A. S. Öğrenci, « Why do students prefer search engines over academic databases », in *2013 12th International Conference on Information Technology Based Higher Education and Training (ITHET)*, oct. 2013, p. 1-3, <https://doi.org/10.1109/ithet.2013.6671054>.
- [24] R. Ihaka et R. Gentleman, « R: A Language for Data Analysis and Graphics », *Journal of Computational and Graphical Statistics*, vol. 5, no 3, p. 299-314, sept. 1996, <https://doi.org/10.1080/10618600.1996.10474713>.
- [25] G. Austin, « Review of The New S Language, A Programming Environment for Data Analysis and Graphics », *The Economic Journal*, vol. 100, no 401, p. 650-652, 1990, doi: 10.2307/2234167.
- [26] C. Severance, « Roy t. fielding: Understanding the rest style », *Computer*, vol. 48, no 6, p. 7-9, 2015.
- [27] L. Pan, M. Xu, L. Xi, et Y. Hao, « Research of livestock farming IoT system based on RESTful web services », in *2016 5th International Conference on Computer Science and Network Technology (ICCSNT)*, déc. 2016, p. 113-116, <https://doi.org/10.1109/iccsnt.2016.8070130>
- [28] Z. Nadia, A. Nadia, et H. Imane, « Towards an Ontology of Dependability Automatic Search Approach with Systematic Mapping », in *Proceedings of the 2Nd International Conference*

on Big Data, Cloud and Applications, New York, NY, USA, 2017, p. 24:1–24:5, <https://doi.org/10.1145/3090354.3090379>.

- [29] F. E. MDARBI, N. AFIFI, I. HILAL, et H. BELHADAOUI, « An Approach for Selecting Cloud Service Adequate to Big Data Case Study: E-health Context », International Journal of Computer Science and Information Security (IJCSIS), Vol. 16, No. 8, August 2018.

## 8 Authors

**Fatima Ezzahra Mdarbi** PhD student at Networks, Computer, Telecom & Multimedia (RITM) Laboratory at the Higher School of Technology at Hassan II University of Casablanca Morocco (UH2C). Doctoral study center engineering science of the National Higher School of Electricity and Mechanics (ENSEM) at UH2C. She is graduate as a computer engineer from the National School of Applied Sciences (ENSA) Hassan I University of Settat Morocco in 2015. Her research interests include Big Data, cloud computing and dependability. She participated in many conferences such as BDCA17, SysCoBioTS19, ICSSD19, and COC19.

**Nadia Zahir** Development engineer. She is graduate from faculty of sciences and technologies, Hassan II University of Casablanca (UH2C). Her area of interest: Ontology and dependability. E-mail: [nadia.zahir.nz@gmail.com](mailto:nadia.zahir.nz@gmail.com)

**Nadia Afffi** is a Professor at Higher School of Technology (ESTC), Hassan II University of Casablanca Morocco (UH2C) since 1993. She received an Electrical Engineering, Computer Systems degree from the Mohammadia School of Engineers (EMI), Mohammed V University - Rabat (1988). She received a PhD in Computer Engineering from UH2C; and HDR in IT in 2014. She is a Senior Member; part of “Software Engineering, Business Intelligence and Security” (SEBIS) research team of “Networks, Computer, Telecom & Multimedia” (RITM) Laboratory (ESTC). She is member of the “Moroccan Association of Information Technology” (AMTI), and member of UH2C Chair on Innovation (UNCHAIN). Her research interests include: Software Architecture; Information Systems Security, Information Systems Modeling, Human Machine Interface, Dependability, Business Intelligence; Cloud Computing Security, Big Data, IoT and Artificial Intelligence. She participated as a consultant then software architect in several national and international projects. She developed several partnerships with international universities, notably in Korea Republic. She coordinated and participated in several international cooperation programs. E-mail: [nafffi@est-uh2c.ac.ma](mailto:nafffi@est-uh2c.ac.ma)

**Imane Hilal** is an Assistant Professor in the School of Information Sciences (ESI) and co-supervises several theses. She holds a Dipl.-Ing. (MSc) from National School of Computer Science and Systems (ENSIAS) in 2009. Then she spent more than eight years in teaching IT development to IT Specialized technician. In 2016 she got a Ph.D. degree from Hassan II University of Casablanca. Her research interests lie in various Computer Science Areas such as: Data Warehousing, Dependability, MDE, Elearning, Big Data, Cloud Computing. She participated in many conferences as Organising Committee or TPC such as: ISI17, MISC18, ArabWIC19 and as Program Committees Chairs in ICSSD19 E-mail: [imanehilal@gmail.com](mailto:imanehilal@gmail.com)

Article submitted 2020-04-21. Resubmitted 2020-05-08. Final acceptance 2020-05-08. Final version published as submitted by the authors.