

Journal of Education and Vocational Research (ISSN 2221-2590)
Vol. 6, No. 4, pp. 24-30, December 2015

Big Data at the Service of Teaching and Scientific Research within the UAE

Karim Aoulad Abdelouarit¹, Boubker Sbihi¹, Noura Aknin²

¹Abdelmalek Essaadi University, Tetuan, Morocco

²Ecole des Sciences de l'Information, Rabat, Maroc
kouladabdelouarit@esi.ac.ma

Abstract: With the emergence of the digital world, data access has become an action as quick and direct that many questions arise about the reliability and value of information. However, the data returned by the online search engines are based on an open and massive environment, which include data of any type, which come from different sources of information. Indeed, in the field of education and especially when using online search, the learner is left with a set of heterogeneous information that does not exist in an orderly format and are not easy to consume. This aspect of data variety represents the second V in the design of the Big Data phenomenon. With the aim to support learners in their search for information, it is proposed to design a specialized structuring tool to combine process and classify the variety of these massive data so they can provide the best result to the learner. The present article deals with the descriptive study of the current state of using online search by UAE's students, and the proposal to explore new methods and approaches through experimentation of our solution for open and massive data environment, in order to enhance learning and scientific research in the UAE.

Keywords: *Big data, e-learning, online search, data structuring, digital pedagogy*

1. Introduction

Search engines provide the user with a large amount of data at each use. However, this mass of heterogeneous and unstructured information at a time, does not always meet the user's expectations against the requirements he expressed. To date, these data have rarely been exploited intelligently with the aim of improving learning and practical information in the academic community, although evidence from other sectors such as marketing, sports, health and technology suggest that the effective use of large volumes of data can provide the education sector the opportunity to improve its systems and results (Macfadyen et al., 2012). In this context, it has proposed to design a tool that can rework the enormous data resulting from online search, making them easily consumable by the learner, and providing relevant and optimal search with consistent and adequate data to expressed needs. The next chapter presents the state of the art concerning the use of online search in Big Data environment by the learner, with discussing the problem of data variety provided by traditional online search engines. Chapter 3 is devoted to the presentation of a new mode of learning that takes as case study the Moroccan University Abdelmalek Essaadi of Tetuan, with the proposal to explore new methods and approaches through the use of our technique of restructuring, classification and presentation of data, tailored to the open and massive data environment, with the aim to enhance learning and scientific research for the UAE's students. The last paragraph presents a general conclusion highlighting a range of perspectives.

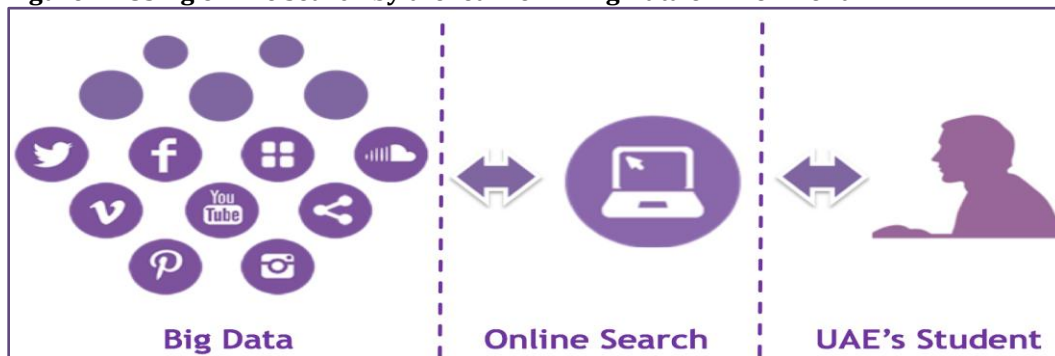
2. The use of online search by the learner in Big Data environment

Definition and state of the art: The significant growth of information in the Internet requires search tools more and more efficient that can discern relevant information from hundreds or even thousands of raw data. However, the quality of the results provided by the search engines is not always relevant, especially when the user comes to compose more than one request (Soussi et al., 2008). Indeed, the online documentation systems and especially the search engines have developed strongly since the advent of Big Data phenomenon. The effectiveness of the search for information, especially on the Web, would be particularly related to the use of the system expertise, including knowledge of the procedures and documentation tools, but also the heuristics and strategies to use in documentation and that, to assess quickly and correctly the content quality and credibility of data and information (Vibert et al., 2007). Today, information is manifested in various

forms: the geo-location, mobile data, data from social networks, video and satellite imagery, customer transactions, motion data of connected objects, etc. This great mass of heterogeneous data has a negative impact for the learner who uses the online search and makes the process of learning or knowledge difficult to succeed. Thus, the use of an online information retrieval support system based on the raw data of Big Data phenomenon has significant strengths. To do this, the need for a fictitious model to represent and process this type of data that are not necessarily text becomes essential. However, semi-structured databases No SQL do not provide enough structure to organize unstructured data, but there are NoSQL engines that implement a variety of data models, and they generally run on flexible storage formats such as JSON. This allows to analyze and to process the data stream in real time (Duggan et al., 2015).

In this context, we mention the solution "DOCUPOLE" which was designed in 2007 to offer an introductory online course in documentary research modeled with MOOC. His new version opened in December 2014, has posted a first significant movement consultation on referencing information (Blondeel et al., 2015). Similarly, we find the system "SARI Onto" which is an online information retrieval support system based on both domain and service ontologies issued from the information search process and document classification (Soussi et al., 2008). Both systems tend towards facilitating access to documental information but without considering the reorganization of heterogeneous data resulting from online research. The Figure 1 shows the appearance of using online search by the learner in the Big Data environment.

Figure 1 : Using online search by the learner in Big Data environment



As shown from this figure, the learner accesses the online search platform for its learning and documentation needs. But finally he left with a multitude of unstructured data and heterogeneous information (text, image, video, ...) that comes from several sources of information (Twitter, Facebook, YouTube, ...). This situation, pushes the learner to question about the value and reliability of the returned information results following the execution of its online search process.

Table1: Some of data types returned by search engines

Data Type	Examples
Web and social media	Click stream data; Twitter feeds; Facebook postings; Web content.
Machine to machine	Utility smart meter readings; RFID readings; Oil rig sensor readings; GPS signals.
Big transaction data	Healthcare claims; Telecommunications call detail records; Utility billing records.
Biometrics	Facial recognition ; Genetics.
Human generated	Call center voice recordings; Mails; Electronic medical records.

Problematic of data variety returned by online search and its impact on pedagogy of teaching: The growing number of available data on Internet makes necessary the use of appropriate tools to assist and support users in their use of online search. Moreover, the techniques generally used by search engines are based on statistical methods and syntactic treatments that not processing the relevance of the user request and the good presentation of the result. Other problems, such as those related to the vocabulary and the data variety, the selection of key words and data filter may be encountered. In this context, helping a user to find the information that he seeks becomes a task increasingly difficult. Hence, the recourse to the inclusion of the semantic level to overcome these difficulties in the information search, this is called the Semantic Web (Berners-Lee et al., 2001). However, the analysis and processing of mass data produced by Big Data phenomenon exceed human capacity, which requires the use of algorithms and powerful computer systems able to explore them. Moreover, the information provided by Big Data is not manifested in an orderly format and are not easy to operate for analysis (Aoulad et al., 2015). This issue concerns us in the context of this article and especially the question of data variety and unstructured information produced by Big Data phenomenon. The Table 1 shows some examples of data types returned by search engines while using online search.

As presented in this table, the information collected with online search engines comes in several types (Web, RFID, and Mail) and from different sources of information. In addition, the emergence of Web 2.0 has made it possible to create and publish any type of content such as courses, exercises, assignments, or bibliographies, and any kind of digital resource that allow an informal collaborative learning known as name of Learning 2.0 (Sbihi et al., 2013). Similarly, Web 2.0 offers several of other collaboration tools such as blogs, wikis, social networks or podcasts and video casts (Sbihi et al., 2013). Consequently, the problematic of data variety exists clearly in the educational field and especially during an online search process used by the learner. At the end of his search, the user is left with a mixture of heterogeneous data that do not always correspond to its requirements and create sometimes a sense of confusion about the reliability and the value of the obtained information.

3. Towards a new learning mode using online search in Big Data environment

Case study of Abdelmalek Essaadi University: The Abdelmalek Essaadi University is one of the four public universities in Morocco; it is considered the primary Northern University. It has consists of 11 institutions, including schools and institutions in the Tangier-Tetuan region distributed at four sites: Tangier, Tetuan, Larache and Martil (El Amrani et al., 2012). It covers three main areas: Science and Technology, Arts and Humanities and Law, Economics and Management. The UAE holds several structural projects and knows a continuous development. It counts with around 41,000 students, 780 teachers, 579 staff members, 35 teams and 29 research laboratories. The University provides students, teachers and administrative partner's universities, his research labs, Internet and other services to enable them to perform their daily work at ease. The university also provides libraries with an important documentary base. For the digital development of the university's work. The UAE has chosen the cloud computing to improve communication, connectivity and collaboration. Following "Google Apps for Education" is the IT solution adopted, it has also reduced IT costs. This suite includes several applications including Gmail (email services), Google Calendar (shared calendaring), Google Docs (documents and presentations), Google Video (10 GB free) and Google Sites (website creation). UAE's students now have the opportunity to work together in real time on the same document while sharing information. At any time, each student will have access to all data through a laptop, tablet or Smartphone connected to the Internet.

As part of this article we want to study the opportunity to experience a new way of learning in the UAE via the proposal to explore new methods and approaches through the use of our technique of restructuring, classification and presentation of data, tailored to the open and massive data environment, with the aim to enhance learning and scientific research within the UAE. This solution can be then added to the cloud computing applications that the University already possesses. The case study is a nominal scenario concerning the use of online search by the learner for his learning or documentation needs on a specific topic or theme. The example illustrated in Figure 2, shows the current state of using the traditional search engine by the UAE's student through its interface to search for information and documentation. Thus, we have chosen the example of searching about the subject of the system "e-learning"(the platform for training and distance learning) as shown in the Figure 2.

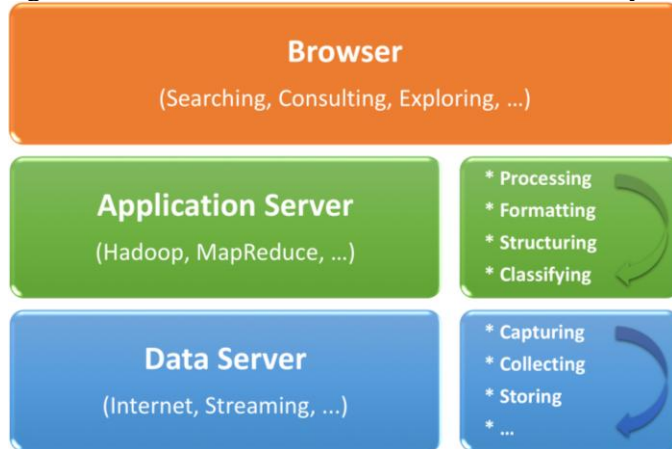
Figure 2: The use of the online search system about the subject "e-learning"



After entering keywords to search the specific theme as presented in the previous figure, the system proceeds to capture all types of data (text, image, video, etc.) about the topic "e-learning" and displays the results in a crude manner and in the same interface of the user, without showing any signs of structuring or insight that can give to the learner a structured view and a prior idea about the result content, which normally can discharge him of an additional phase of data discovery by browsing each result separately to check finally whether it corresponds to its requirements or not.

Towards a based tool on Big Data to improve learning and research in the UAE: It is important to identify a sophisticated strategy to combine different types of data in a way that they provide the best result to the learner. In this context, we propose to develop a technique based on a tool that integrates all structured and unstructured data in one data layer to facilitate access and optimize relevance with search adequate and consistent results according to the expectations of the learner (Aoulad et al., 2015). The Figure 3 shows the functional architecture of the future system.

Figure 3: The functional architecture of the future system



As shown in the figure, the job consists in to develop a top layer to the layer of raw and massive data provided by Big Data, which will take charge of dealing with the large and unstructured data and subsequently offer

storage, research and dissemination of consistent information with the data requested by the learner through the online search. To do this, it is proposed to capture initially the metadata values of the meaningful information constituting the results of the gross search returned by traditional search engines, and then organize them in the form of well-formatted structures to present them to the learner as an overview or summary of the result contents returned by the online search. This will allow it to browse the essential and meaningful information, allowing it to decide on the follow mining them or not. This technique will discharge the learner the phase concerning the random exploration of the raw results returned by the classic online search, as these data do not always meet his expectations and take unnecessarily time consuming for their consultation. Technically speaking, the task consists to organize the resulting data from the online search as data overviews that summarizes their contents. Thus, we present in the following Figure 4, the Java class used to organize information about it by choosing the relevant and corresponding metadata to the proper presentation of search results:

Figure 4: The Java class used to organize the search results

```

Public class Search Result {

    private String title;           // The title of the search result
    private String type;           // The type of the search
                                   result (post, image, video)
    private Calendar publication;  // The publication date of the
                                   search result
    private String preview;       // The text preview for the
                                   content of the search result
    private String thumbnail;     // The path of the image preview
                                   for the content of the search
                                   result if it's exists
    private String source;       // The content source of the
                                   search result (Wikipedia,
                                   Twitter, YouTube, ...)
}

```

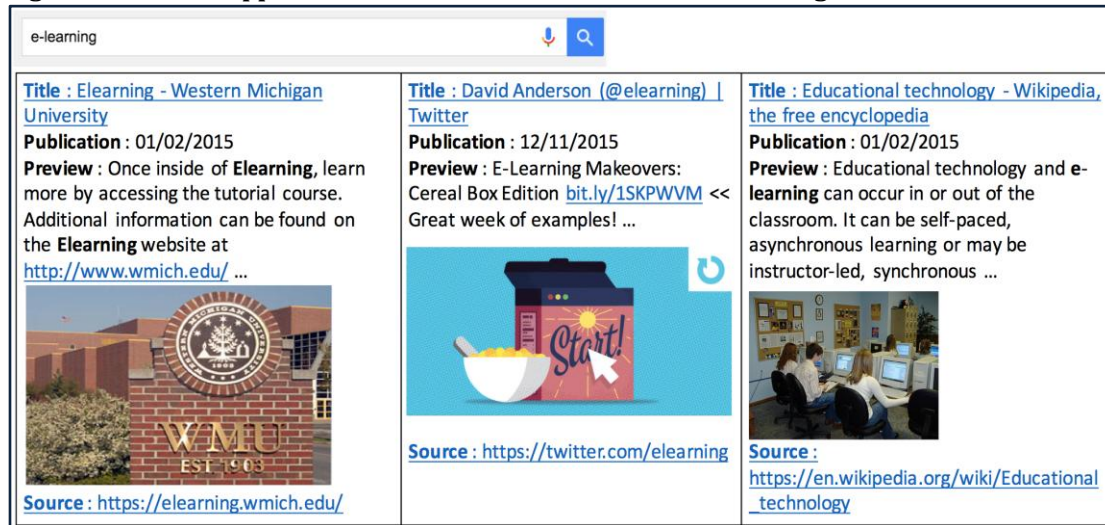
The "Search Result" class refers to the summary of the content to be presented for the user in his first search result screen. So, we can find the information about the title of the result, his type (text, image, video, ...), his date of publication (allows to order the result by the most recent element), a preview of the content (a kind of descriptive text), an illustration image if it exists and finally the source of the information (Wikipedia, Twitter, YouTube ...). Thus, the nominal usage scenario of online search that we have already mentioned in the case study of the UAE paragraph at the top, can be realized through this new system as described at the Figure 5.

Figure 5: Towards a new structuring technique of the online search results



Thus, the system user (learner) accesses the search interface to enter the keywords of his information request. These keywords are sent to the system to retrieve the corresponding information to their semantics from the Big Data layer. Then, the captured data is processed at the middle layer that concerns our future system that it's responsible for structuring the processed data, classifying and sending them at the end to the results page of the system, which will be displayed to the learner in an organized and well presented manner. It then finds the data related to the type "post" classified in the same category even though they come from different sources (Facebook, Twitter, Google+, ...), also for image data type (Instagram, Picasa, ...), video data type (YouTube, Vimeo, Dailymotion, ...), and finally the audio data type (Soundcloud, ...). In order to have a clearer idea about the future system, the Figure 6 exposes the real appearance of the final user interface that performs the online search process executed by the learner.

Figure 6: The real appearance of the final user interface concerning the future solution



As shown, the search results will be organized in a way to present to the learner an overview or summary data of the results contents. Thus, the learner can now to have a prior idea about the rest of content and to decide to continue exploring the significant result or not depending their needs. In addition, this new process will allow him to economize his research time through browsing only the relevant search results and then exploring the most adequate to its needs.

4. Conclusion and future work

The integration and structuring of various data types is a rare feature that we can find on the big data solutions in the current market. That is why we focused our future solution on the design of a technique that integrates the set of structured and unstructured data in a single data layer to facilitate access and to offer optimum ergonomic presentation of online search results achieved by the learner. Finally, students from the UAE will run online search process for their learning or documentation needs on a given theme, with flexible and mastered manner. Thus, they can browse faster all results of their search and access only to appropriate and consistent results that meet better their requirements. Moreover, this technique will relieve them of unnecessary exploration phase of the non-corresponding results in their quest for information and will save their search time. The solution is to detect language elements and turn them into a kind of summary data that can be easily operated and consumed by the user. The adopted method consist first to the capture of metadata elements and to classify its information according to each type of result returned, providing a list of significant results and presented in an ergonomist way. As a perspective of this work, we will conduct a study survey of criteria and factors affecting learner's environment against massive data offered by Big Data, and through the study of use cases Search online by the learner for a learning purpose or documentation on a given theme. This, through the creation and submission of the corresponding questionnaires to a sample of the UAE students using online research for their learning. This will be followed by a detailed analysis of the results collected from these questionnaires and that will frame the functional and technical requirements of

the future solution and begin the development of the corresponding tool based on a fictitious model to process the massive and heterogeneous information provided by Big Data.

References

- Aoulad, A. K., Sbihi, B. & Aknin, N. (2015). Big-Learn: Towards a tool based on Big Data to improve research in an e-learning environment. *International Journal of Advanced Computer Science and Applications*, 6(10).
- Berners-Lee, T., Hendler, J. & Lassila, O. (2001). The semantic web. *Scientific American*, 284(5), 28-37.
- Blondeel, S., Pochet, B., Collette, C., Delbushaye, T., Frédéric, F., Lanners, C. & Thirion, P. (2015). DOCUPOLE: un cours en ligne d'initiation à la recherche documentaire modélisé en MOOC. In *Former aux compétences informationnelles à l'heure du Web 2.0 et des discovery tools* (pp. 20-28).
- Duggan, J., Elmore, A. J., Stonebraker, M., Balazinska, M., Howe, B., Kepner, J. & Zdonik, S. (2015). The Big DAWG Polystore System. *ACM SIGMOD Record*, 44(2), 11-16.
- El Amrani, C., Tani, H. G. & Eloutouate, L. (2012). Towards Setting up IBM Cloud Computing VCL at Abdelmalek Essaadi University. *International Journal of Computer Applications*, 56 (16), 26-28.
- Macfadyen, L. P., Dawson, S., Pardo, A. & Gasevic, D. (2014). Embracing big data in complex educational systems: The learning analytics imperative and the policy challenge. *Research & Practice in Assessment*, 9(2), 17-28.
- Sbihi, B., El Kadiri, K. & Aknin, N. (2013). Towards an Implementation of the Concepts of E-Learning 2.5 through one Group of ten Master's Learners: Case of the UML Course. *International Journal of Emerging Technologies in Learning (ijET)*, 8(4), 68-73.
- Soussi, R., Mustapha, N. B., Zghal, H. B. & Afaure, M. A. (2008). Un système d'aide à la recherche d'informations en ligne basé sur les ontologies. In *CORIA 2008* (pp. 483-490).
- Vibert, N., Rouet, J. F., Ros, C., Ramond, M. & Deshoullieres, B. (2007). The use of online electronic information resources in scientific research: The case of neuroscience. *Library & Information Science Research*, 29(4), 508-532.