

ULearn: Personalized Medical Learning on the Web for Patient Empowerment

Marco Alfano^{1,5}[0000-0001-7200-9547], Biagio Lenzitti²[0000-0003-2664-7788], Davide Taibi³[0000-0002-0785-6771], Markus Helfert⁴[0000-0001-6546-6408]

¹ Lero, Dublin City University, Dublin, Ireland

² Dipartimento di Matematica e Informatica, Università di Palermo, Palermo, Italy

³ Istituto per le Tecnologie Didattiche, Consiglio Nazionale delle Ricerche, Palermo, Italy

⁴ Lero, Maynooth University, Maynooth, Co. Kildare, Ireland

⁵ Anghelos Centro Studi sulla Comunicazione, Palermo, Italy

marco.alfano@lero.ie, biagio.lenzitti@unipa.it, davide.taibi@itd.cnr.it,
markus.helfert@lero.ie

Abstract. Health literacy constitutes an important step towards patient empowerment and the Web is presently the biggest repository of medical information and, thus, the biggest medical resource to be used in the learning process. However, at present, web medical information is mainly accessed through generic search engines that do not take into account the user specific needs and starting knowledge and so they are not able to support learning activities tailored to the specific user requirements. This work presents “ULearn” a meta engine that supports access, understanding and learning on the Web in the medical domain based on specific user requirements and knowledge levels towards what we call “balanced learning”. Balanced learning allows users to perform learning activities based on specific user requirements (understanding, deepening, widening and exploring) towards his/her empowerment. We have designed and developed ULearn to suggest search keywords correlated to the different user requirements and we have carried out some preliminary experiments to evaluate the effectiveness of the provided information.

Keywords: Patient empowerment, search as learning, e-health, Health literacy, Health seeking behavior.

1 Introduction

“Engaging and empowering people & communities” constitutes the first of the five strategies of the “Framework on integrated people-centred health services” of the World Health Organization (WHO). It calls for a fundamental shift in the way health services are funded, managed and delivered and presents a compelling vision of a future in which all people have access to health services that are coordinated around their needs, respects their preferences, and are safe, effective, timely, affordable, and of acceptable quality [1]. In this perspective, the way in which people access, understand and learn about health information plays a key role in enabling patient empowerment [2][3].

When it comes to the access of online information, search engines are more and more used as a tool to support users in finding information in the non-homogeneous and continuously evolving World Wide Web. In particular, when users look for information on a new topic, related either to health or other subjects, generic search engines, such as Google™ or Bing™, are undoubtedly the most used tools as a starting point. However, the amount of information published on the Web has reached a huge dimension and the number of search results can be potentially enormous so leading users to spend a lot of time in searching what they are looking for, even with the risk of not finding the expected information. This is worsened by the fact that the majority of users take into account only the first results retrieved by the search engines (at most the first 20-30 results). Consequently, those results are crucial and play a key role in influencing users understanding in a specific domain, even though might not contain all the real “useful” information.

At the same time, the connection among search processes, learning and knowledge building is emphasized in several works in literature [4], [5]. This is bringing, as a natural consequence, to the development of a new research field which connects the educational sphere with information seeking. In particular, “search as learning” investigates the relationship between searching and learning where “the information seeking is conceptualized as a learning process, and learning as an outcome of the information seeking process” [6].

Generic search engines have not been designed to support learning processes, therefore it is necessary to develop more suitable tools that, on one hand offer the possibility to search for information on the whole Web, as the generic search engines do, and, on the other hand, are able to support learning activities based on effective searching processes [7][8].

This paper, based on a previous work presented in [9], aims to evolve the idea of using a meta engine to create knowledge paths on the web. In particular, we present “ULearn”, a new version of the meta engine that supports learners in achieving the objective of a “*balanced learning*” through personalized learning paths in the medical domain, based on specific user requirements. To this end, we consider four different learner categories: “basic”, “deep”, “wide” and “explorer”, each of them with different objectives and needs. ULearn suggests words correlated to each of those learner categories thus supporting the creation of medical learning activities tailored to the real user needs.

The paper is organized as follows. Section 2 describes the basic principles upon which ULearn is based with particular respect to the “balanced learning” approach. Section 3 presents the software architecture of ULearn and Section 4 describes its implementation and preliminary experimental results. The final section presents some conclusions and ongoing work.

2 **Balanced Learning Methodology**

The research field of “search as learning”, recently developed, highlights the connections between the information seeking and the learning processes [4], [5]. Users, generally, have different needs when performing a search especially for knowing and

learning, and that is particularly emphasized when users want to shape their understanding on a new topic in the medical domain.

In what follows we distinguish "learning searchers" from "focused searchers": a "learning searcher" is not looking for a specific information but explores and navigates the web to increase his/her knowledge (e.g., a user who wants to learn more about *diabetes*), a "focused searcher", instead, is looking for a specific piece of information and uses a general-purpose search engine to find it (e.g., a hospital that deals with a specific disease). For what concerns the learning process for patient empowerment, the "learning searcher" is the category directly involved and it is the main target of this study.

With regards to the "learning searchers", we consider four main categories for four different learning needs. They have been identified starting from the categorizations presented in [9], [10], [11]:

- a "basic learner" who knows little about a medical topic and desires to learn the fundamental aspects of the topic, i.e., looks for medical information strictly correlated to the searched keyword(s);
- a "deep learner" who wants to learn more specific details on a medical topic he/she already knows, i.e., looks for medical information that provides the details of the searched keyword(s);
- a "wide learner" is not so interested in focusing on the details of a medical topic but rather prefers to expand his/her medical knowledge domain with topics that are partially related to the starting topic;
- an "explorer" who considers the initial keyword(s) just as a starting point and wants to expand his/her medical knowledge domain with topics that are loosely related to the initial keyword(s).

Each searcher category has different objectives influencing the learning process. However, since patient empowerment calls for autonomy from the user, it is he/she who will decide whether, when and how to cover the different learning aspects, in order to reach what we call a "balanced learning". Fig. 1 presents the pyramid that leads to the "balanced learning" where the four sides of the pyramid (Fig. 1 only shows two sides of the pyramid for graphical simplicity) correspond to the four learning needs, related to basic, deepening, widening and exploring categories. Fig. 1 also shows the knowledge already acquired on each specific dimension and, consequently, the learning gap to the pyramid top and, then, to the balanced learning.

Even though, a balanced learning, in terms of a learning process that evenly covers the four different categories, is desirable, it is the user who, ultimately, establishes his/her learning goals for achieving empowerment and what are the learning needs (if any) on each side of the pyramid to reach a "balanced learning". Thus, the top of the pyramid is not reached when specific knowledge levels are reached in each side of the pyramid but, rather, when the user has satisfied all his/her learning needs with the learning activities in the four categories.

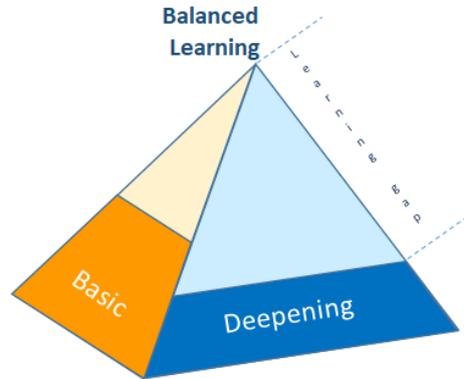


Fig. 1. Learning pyramid.

3 ULearn Meta Engine

ULearn is a meta engine that implements the “search as learning” methodology presented above. It allows a user to specify one or more keywords and provides words for each of the four categories presented above. Fig. 2 shows its basic architecture.

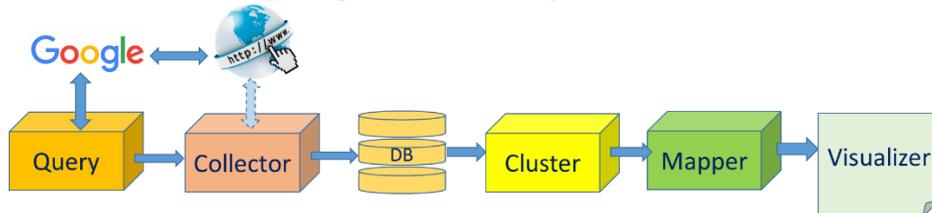


Fig. 2. ULearn architecture.

The “QUERY” module takes the keyword(s) specified by the user and the number n of documents (web pages) to be analyzed. It then uses Google™ to search the World Wide Web with this keyword(s) and takes the first n results creating a collection of n links to the related web pages.

For the found links, the “COLLECTOR” module retrieves the related web pages, cleans them (by removing tags and common words) and stores the remaining m words in the “DB” database together with the total number of pages that contain the word and the number of occurrences of a word in each web page divided by the total number of words in the page (term frequency - tf).

The “CLUSTER” module applies the K-means algorithm to a bidimensional matrix made up, for each word, by the normalized number of web pages containing the word (document frequency - df) and the normalized term frequency. The K-means algorithm creates four clusters, related to the four categories, starting from the outermost corners of the normalized matrix, i.e., $C_0(0,0)$, $C_1(1,0)$, $C_2(0,1)$ and $C_3(1,1)$.

The “MAPPER” module associates the clusters to the four learner categories, presented above, considering that each word has a specific correlation with the initial

keyword(s) and will allow a specific type of navigation. Thus, the “understanding” words are likely to be conceptually close to the initial keyword(s). They will be used by a basic learner for an understanding and learning of the related knowledge domain (e.g., "insulin" and "type" for the *diabetes* keyword). The “deepening” words are terms that are likely to have a strong correlation to the initial keyword(s) (in terms of number of occurrences) but only appear in a few documents so they are likely to represent specific topics inside the semantic domain (e.g., "periodontal" and "diabetesvoice" - the latter is the online magazine of the International Diabetes Federation - for the *diabetes* keyword). The “widening” words are terms that are likely to have a loose correlation (in terms of number of occurrences) to the initial keyword(s) but appear in many documents so they are likely to represent topics at the border of the of the semantic domain (e.g., "obesity" and "exercise" for the *diabetes* keyword). Finally, the “exploring” words are terms that are likely to have a very loose correlation (both in terms of number of occurrences and pages) to the initial keyword(s) so they are likely to represent topics that allow the user to easily explore other semantic domains allowing him/her to increase the knowledge of those domains may be finding “something” even more interesting of the initial search thanks to a serendipitous discovery [12] (e.g., "hyperphagia" and "periodontist" for the *diabetes* keyword).

Thus, starting from the results of previous studies [9], [10], [11], we assume the following rules:

- a word that appears in many pages with many occurrences can be used for “understanding” and “basic learning”;
- a word that appears in a few pages with many occurrences can be used for “deepening” the knowledge;
- a word that appears in many pages with a few occurrences can be used for “widening” the knowledge.
- a word that appears in a few pages with a few occurrences can be used for “exploring” the knowledge.

And, consequently, the following associations apply:

- the C_3 cluster is associated to the “understanding” category;
- the C_1 cluster is associated to the “deepening” category;
- the C_2 cluster is associated to the “widening” category;
- the C_0 cluster is associated to the “exploring” category.

The "VISUALIZER" module will present the user with a maximum number of m words for each category (where m has been presently set equal to five for the readability of the results but can be easily changed).

4 ULearn Implementation and Experimental Results

We have implemented the ULearn meta engine following the architecture presented in the previous section and using the PHP language and MySQL database. Fig. 3 shows the input page that allows the user to specify the keyword(s) to be searched, the learning category (understanding, deepening, widening and exploring) and the number of pages to be analyzed.

We have run some experiments using six different medical keywords. Three terms are popular medical terms i.e., *Arthritis*, *Diabetes* and *Hepatitis*. The other three have been

chosen among the most complex medical terms¹, i.e., *Aphthous Stomatitis*, *Bradykinesia* and *Epistaxis*.

For each term, we have used the QUERY module to take the first fifty Google results - going beyond the number of twenty-three results manually analysed by a user (as seen above) - and create a set of fifty links to web pages. Then, we used the COLLECTOR to retrieve the web pages, extract the words, make the computation presented in the previous section and store the results in the DB.

The next step was to run the K-means algorithm of the ANALYZER to each bidimensional matrix. Its application poses some limitations to be taken into account as described in [13]. They are the possible indeterminacy of the exact number of clusters, the potential different results in presence of outliers, the non-optimal results in presence of not well distributed data and, finally, the possible presence of empty clusters. Starting from the assumption that the aggregations sought were for the four different learning needs, the keywords of the research were excluded from the data in order to avoid outliers. Moreover, we chose, as initial centroids, the external points to ensure an initial clustering with at least one value per cluster and avoid empty clusters.

Fig. 3. ULearn input page.

Table 1 presents the results summary of the clustering algorithm for the six keywords. In particular, for each cluster it shows the X and Y coordinates of the centre and the number of word occurrences.

Table 1. Summary of results of the clustering algorithm.

Keyword	C ₀	C ₁	C ₂	C ₃
<i>Arthritis</i>	X = 0.042305 Y = 0.007535 No = 10803	X = 0.047727 Y = 0.444596 No = 13	X = 0.280934 Y = 0.00972 No = 593	X = 0.29404 Y = 0.012189 No = 344
<i>Diabetes</i>	X = 0.032328 Y = 0.015297	X = 0.025625 Y = 0.418576	X = 0.405283 Y = 0.033976	X = 0.786458 Y = 0.321185

¹<https://www.prdaily.com/17-complicated-medical-terms-and-their-simpler-explanations/>

	No = 13933	No = 59	No = 375	No = 21
<i>Hepatitis</i>	X =0.041577 Y =0.011861 No = 11117	X = 0.031586 Y =0.3524 No = 63	X =0.364729 Y =0.022705 No = 521	X =0.725 Y =0.145769 No = 60
<i>Aphthous Stomatitis</i>	X = 0.054744 Y =0.011107 No = 8263	X = 0.029457 Y =0.316276 No = 37	X = 0.382242 Y =0.0127 No = 679	X =0.674419 Y =0.162734 No = 68
<i>Bradykinesia</i>	X = 0.047806 Y =0.010522 No = 8927	X =0.018803 Y =0.483662 No = 10	X = 0.326734 Y =0.011354 No = 548	X =0.676374 Y =0.033181 No = 60
<i>Epistaxis</i>	X = 0.044088 Y =0.057612 No = 8176	X = 0.026454 Y =0.350072 No = 461	X = 0.281967 Y =0.047219 No = 722	X = 0.664286 Y =0.063755 No = 92

The exam of Table 1 shows that the centres of the clusters are approximately on the same positions. The C_0 cluster presents a much higher number of words compared to C_2 (second cluster in terms of number of words) and C_1, C_3 . The graphical representation of the four cluster for the *Diabetes* keyword is presented in Fig. 4. The blue dots (bottom-left) belong to C_0 , the red dots (top-left) belong to C_1 , the yellow dots (centre) belong to C_2 and the green dots (right) belong to C_3 .

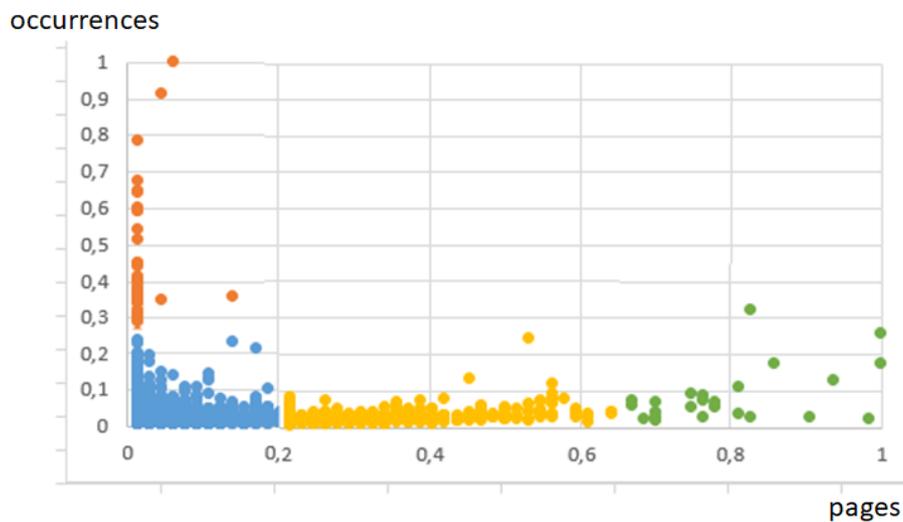


Fig. 4. The four clusters for the *Diabetes* keyword.

Considering the association between cluster and learner category presented in the previous section, Table 2 shows, for each keyword and for each category, the first five words of that category.

Table 2. ULearn results for the six medical keywords.

Keyword	Understanding	Deepening	Widening	Exploring
<i>Arthritis</i>	1. joint 2. joints 3. osteoarthritis 4. rheumatoid 5. disease	1. lilia 2. proxima 3. ireland 4. organization 5. cleveland	1. original 2. foods 3. kids 4. jia 5. acetaminophen	1. express 2. verywell 3. salute 4. assets 5. emmerdale
<i>Diabetes</i>	1. insulin 2. type 3. health 4. glucose 5. disease	1. periodontal 2. diabetesvoice 3. melton 4. wdf 5. wbur	1. obesity 2. exercise 3. diabetic 4. diet 5. screening	1. hyperphagia 2. periodontist 3. symlinpen 4. flexpen 5. summaries
<i>Hepatitis</i>	1. virus 2. infection 3. viral 4. chronic 5. disease	1. ethanol 2. nsw 3. zealand 4. parenthood 5. fibroscans	1. viruses 2. hcv 3. hav 4. donor 5. transplantation	1. reg 2. hepatovirus 3. nys 4. fund 5. credit
<i>Aphthous Stomatitis</i>	1. oral 2. ulcers 3. stomatitis 4. recurrent 5. ulceration	1. wiley 2. ibd 3. sage 4. aocd 5. idcsu	1. pubmed 2. med 3. edit 4. data 5. rau	1. osteopathic 2. columns 3. clinic 4. irbesartan 5. osmosis
<i>Bradykinesia</i>	1. disease 2. parkinson 3. movement 4. patients 5. parkinsons	1. wiley 2. smiling 3. aps 4. posed 5. dictionary	1. scholar 2. google 3. training 4. cancer 5. subjects	1. merriam 2. nursing 3. appendicular 4. xplore 5. univ
<i>Epistaxis</i>	1. health 2. medications 3. bleeding 4. rarely 5. nosebleed	1. institution 2. faction 3. coli 4. vulgaris 5. escherichia	1. youtube 2. video 3. financial 4. advice 5. appointment	1. checker 2. instagram 3. courses 4. committees 5. clogged

The ULearn meta engine supports search as learning processes by leveraging these results. Starting from an initial search keyword a learner can use the meta engine to obtain other keywords to be used in a subsequent search as learning activity according to his/her learning needs. As an example, assuming that a learner starts a search as learning activity with the word *Diabetes*, the ULearn meta-engine will return specific keywords to be used in further search activities depending on which categories

(understanding, deepening, widening or exploring) the user is interested to expand for his/her empowerment. Therefore, for users who are interested in a basic understanding, ULearn suggests as related keywords: insulin, type, health, glucose, disease, that are popular terms connected to the Diabetes topic. For whom interested in widening the knowledge in the Diabetes topic ULearn will suggest to continue the online searching by using keywords such as: obesity, exercise, diabetic, diet, screening. In this perspective, the ULearn system will contribute to promote personal learning experiences based on search as learning processes (aimed at achieving learner empowerment) on the four faces of the pyramid towards the achievement of a “balanced learning”.

The results presented above are quite satisfying mainly for the Understanding/Basic Learning category because they provide some basic keywords that allow a user to understand the meaning of the keyword and the related elements. The same applies to the Exploring category because the words are quite unrelated and allow an easy exploration of other subject domains. For what concerns the Deepening and Widening categories there is a mix of medical words that actually help in deepening and widening the medical knowledge domain and words that appear less correlated to the medical context.

As a further step, we have filtered the words, only taking the medical words, by means of the system presented in [2] for simplifying medical terminology and that uses the Unified Medical Language System (UMLS) metathesaurus that has been developed by the US National Library of Medicine and contains the most comprehensive collection of medical vocabulary [14]. By filtering the medical words, we noticed that “Understanding” and “Deepening” clusters got quite close as well as the “Widening” and “Exploring” clusters. This is also explained by the fact that the search now is more focused on the medical domain. As a consequence, we have grouped the four clusters to two clusters, namely the “Understanding/Deepening” cluster and the “Widening/Exploring” cluster. The first five medical words of these new clusters are reported in Table 3.

Table 3. ULearn medical results for the six medical keywords.

Keyword	Understanding/Deepening	Widening/Exploring
Arthritis	<ol style="list-style-type: none"> 1. joint 2. disease 3. health 4. margin 5. research 	<ol style="list-style-type: none"> 1. screening 2. cells 3. national 4. kidney 5. training
Diabetes	<ol style="list-style-type: none"> 1. type 2. health 3. disease 4. risk 5. medical 	<ol style="list-style-type: none"> 1. obesity 2. exercise 3. diabetic 4. diet 5. screening
Hepatitis	<ol style="list-style-type: none"> 1. virus 2. disease 3. health 	<ol style="list-style-type: none"> 1. diseases 2. dna 3. rna

	4. risk 5. acute	4. cells 5. infectious
Aphthous Stomatitis	1. ras 2. disease 3. mucosa 4. diseases 5. medical	1. cells 2. herpes 3. virus 4. professional 5. drug
Bradykinesia	1. disease 2. parkinson 3. medical 4. dopamine 5. health	1. cancer 2. resources 3. cells 4. data 5. gene
Epistaxis	1. health 2. support 3. medical 4. disorders 5. sinus	1. video 2. monitoring 3. community 4. oncology 5. willebrand

The results shown in Table 2 are used by the ULearn meta engine to support a learning process in the four categories and in other subject domains. The results shown in Table 3 allow ULearn to support a learning process in two different categories focusing more on the medical field. Although, more experiments are needed, we can preliminary assume to use the four categories for non-medical experts, who have broader needs in terms of learning, and the two categories for medical experts, who have more specific/narrow needs in terms of learning, as discussed in [3], [15]. In this case, we can assume that the balanced learning occurs when the experts mainly progress through the “deepening” and “widening” learning dimensions.

5 Conclusions and Future Work

This paper has presented ULearn, a new meta engine that allows a personalized medical learning process for patient empowerment based on specific user requirements. We have designed and implemented a prototype and carried out some preliminary experiments to cluster search keywords to be used for a search as learning activity.

The preliminary experimental results are satisfying but more experiments are needed to better identify the user needs and the starting knowledge level (e.g., medical expert or non-expert) so to provide the search keywords to the user and allow him/her to progress in his/her learning activity accordingly.

As a future work, we plan to use semantic analysis to overcome the limit of text elaboration (connected for instance to synonyms and acronyms) and leveraging new approaches based on interconnected knowledge graphs. Moreover, the keyword provided by ULearn can be connected to correlated web pages (or parts of them) in order to provide users not only with keywords but also with whole learning contents.

Finally, even though the k-means is one of the most popular clustering algorithms [16] used in the fields of information retrieval, computer vision and pattern recognition, other clustering techniques will be investigated to improve the results.

Acknowledgements

This work was partially supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 754489 and by Science Foundation Ireland grant 13/RC/2094 with a co-fund of the European Regional Development Fund through the Southern & Eastern Regional Operational Programme to Lero - the Irish Software Research Centre (www.lero.ie).

References

1. World Health Organization. (2016). Framework on integrated, people-centred health services: Report by the Secretariat. World Health Assembly, (A69/39), 1–12.
2. Alfano M., Lenzitti B., Lo Bosco G., and Taibi D. (2018). Development and Practical Use of a Medical Vocabulary-Thesaurus-Dictionary for Patient Empowerment. In Proceedings of the ACM 19th International Conference on Computer Systems and Technologies (CompSysTech'18), Boris Rachev and Angel Smrikarov (Eds.). ACM, New York, NY, USA, 88-93. DOI: <https://doi.org/10.1145/3274005.3274017>
3. Alfano M. Lenzitti B., Taibi D. and Helfert M. (2019). Provision of Tailored Health Information for Patient Empowerment: An Initial Study. In Proceedings of the ACM International Conference on Computer Systems and Technologies (CompSysTech'19).
4. Eickhoff, C., Gwizdka, J., Hauff, C. et al. (2017). Introduction to the special issue on search as learning. *Inf Retrieval J.* <https://doi.org/10.1007/s10791-017-9315-9>
5. Ghosh S., Rath M., and Shah C. (2018). Searching as Learning: Exploring Search Behavior and Learning Outcomes in Learning-related Tasks. In Proceedings of the 2018 Conference on Human Information Interaction & Retrieval (CHIIR '18). ACM, New York, NY, USA, 22-31. DOI: <https://doi.org/10.1145/3176349.3176386>
6. Barker J. and Kupersmith J. Recommended Search Strategy: Analyze your topic & Search with peripheral vision. (2009). <http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Strategies.html>
7. Taibi, D., Rogers, R., Marenzi, I., Nejd, W., Ahmad, Q. A. I., & Fulantelli, G. (2016). Search as research practices on the web: the SaR-Web platform for cross-language engine results analysis. In Proceedings of the 8th ACM Conference on Web Science.
8. Fulantelli, G., Marenzi, I., Ahmad, Q. A. I., & Taibi, D. (2016). SaR-Web-A Tool to Support Search as Learning Processes. In SAL@ SIGIR.
9. Alfano M. and Lenzitti B. (2010). U-Search: A meta engine for creation of knowledge paths on the web. In Proceedings of the ACM International Conference on Computer Systems and Technologies (CompSysTech'2010). ACM, New York, NY, USA, 442-447. DOI: <https://doi.org/10.1145/1839379.1839457>
10. Alfano M., Lenzitti B and Lo Bosco G. (2015). U-MedSearch: A Meta Search Engine of Medical Content for Different Users and Learning Needs. In Proceedings of the International Conference on e-Learning (e-Learning'15).
11. Pang, P. C. I., Verspoor, K., Chang, S., & Pearce, J. (2015). Conceptualising health information seeking behaviours and exploratory search: result of a qualitative study. *Health and Technology*, 5(1), 45–55. <https://doi.org/10.1007/s12553-015-0096-0>
12. Campos J. and Dias de Figueiredo A. Searching the Unsearchable: Inducing Serendipitous Insights. Proceedings of the Fourth International Conference on Case-Based Reasoning. Vancouver, Canada, 2001.

13. Shraddha S. and Naganna S. (2014). A review on K-means data clustering approach. *International Journal of Information & Computation Technology*. ISSN 0974-2239 Volume 4, Number 17 (2014), pp. 1847-1860. International Research Publications House.
14. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res.* 2004;32(Database issue):D267–D270. doi:10.1093/nar/gkh061
15. Alfano M. Lenzitti B., Taibi D. and Helfert M. (2019). Facilitating access to health web pages with different language complexity levels. In *Proceedings. of the 5th International Conference on Information and Communication Technologies for Ageing Well and e-Health (ICT4AWE 2019)*, pp. 113-123, 2-4 May 2019, Heraklion-Crete.
16. Anil K.Jain. (2010). Data clustering: 50 years beyond K-means, *Pattern Recognition Letters* Volume 31, Issue 8, 1 June 2010, Pages 651-666