Inducing Academic Databases Direct User over Popular Search Engines

Rafael Ibarra and Efrain Avila

Dirección General de Bibliotecas, UNAM, Mexico

Abstract: Proposes to directly prompt reputed database use over search engines by means of a 3D self- test method and a Linguistic Storm. A 3D self-test schema is used to make users aware of the existing databases within their own libraries pointing to exploit them. The 3D schema consists of three graded axes that involve: a) users´ academic level; b) electronic resources; and, c) linguistic steps to make the IR results pertinent, not repeated (relevant). Additionally, a brief didactic procedure to profit the library's collections is presented. Furthermore, some editors (Proquest, Elsevier and EBSCO) approaches to include their collections references into the popular search engine Google, are discussed. Finally, some remarks on the didactic procedure are included.

Keywords: Information Retrieval, Linguistic Storm, Information Literacy, Linguistics, Didactics, Google.

1. Introduction

In this second decade of the millennium search engines have become numerous scholar users' first option for information retrieval, in such a way that some editors of re-known scientific sources have relinquished their collections to be searchable through popular search engines like Google, or Google Scholar. Most new users are certainly stunned with search engines´ results and, regularly, ignore their libraries´ collections. Furthermore, subscribed collections editors are rising the exponential growth of documents, not only of paralysing results, but alarming frustration. How will users, teachers and librarians face this underlying framework?

For example, ProQuest is enabling the full text of its scholarly journal content to be indexed in Google Scholar, improving discovery and research outcomes.[1]

---

[1] Why ProQuest is working with Google to improve research workflows.

By his part, Ale de Vries (2015), the Science Direct (SD) product manager, from Elsevier, declared:

> *About Google/Google Scholar: we're making good progress. As you may be aware, we did a pilot with some journals on SD first, and now we are working to get them all indexed. We're making good progress there – it's a lot of content to be crawled, but going along nicely. Both Google Scholar and main Google are gradually covering more and more of our journals.[2]*
> *EBSCO has streamlined the process for users to move from a search result in Google or Google Scholar to the corresponding item in an EBSCO host database by providing a combination of IP- addresses and permanent links to the available resource.[3]*

In contrast, tutors, teachers and librarians have been striving to make/establish a clear difference between documents obtained from search engines and the ones gotten from academic databases. In the experts academic fields it is known that, among several differences, documents coming from popular search engines lack reliability because of their *biased* nature.

For example, users can hardly understand why a document found in popular search engines, with the same title, the same author and the same year, can be different from the one found in an academic data base. They overlooked the publishing process of peer review that may demand relevant changes from the authors writing.

For instance, let us suppose that a committee of a specific field of knowledge published the proceedings of their annual congress, but one of the authors is called to publish his/her contributing paper in an indexed journal. We know that in the publishing process there must be a review made by peers to save probable errors, lack of focus, or requires a better management on language and images, figures, tables and so on. The differences, among several others, are precisely these kind of aspects. The final paper is generally improved in an indexed journal than the paper published in proceedings, with the same title, the same author and, probably, the same year.

As a consequence of these three editors' decisions, it seems to be that academic heads' efforts must be a lot more hardworking to connect users and their library

---

http://www.proquest.com/blog/pqblog/2015/Why-ProQuest-is-working-with-Google.html 24, March, 2015. Accessed March, 8, 2016.

[2] Science Direct-ly into Google. http://toc.oreilly.com/2007/07/science-directly-into-google.html 3, July, 2007. Accessed March, 8, 2016

[3] http://support.ebsco.com/knowledge_base/detail.php?id=3590 January, 2016. Accessed March, 8, 2016.

services. However, the core of the users' frustrations does not end by googling their information necessities, since it is not the virtual space what impedes good results, but the use of words that users frequently miss when consulting academic information. Users' linguistic limits are certainly not new problems to get the information they need, and they were detected and interestingly illustrated more than 40 years ago.

In an expensive and thorough effort of 225 information agencies, the National Technical Information Service of the United States tried to build up a unique thesaurus, from four different ones: Defense Documentation Center, Nasa, Subject headings used by the Atomic Energy Commission and the TEST (Thesaurus of Engineering and Scientific Terms) to make retrieval in the easiest possible way. The result was far from aceptable: Urbach (1973) *there remain inconsistencies in indexing and complicates retrieval.*[4]

As an approach to a solution, the report discloses: Urbach (1973):

> *"… the end user will simply have to learn to handle the additional vocabularies as he searches the file. This is not a solution that will enhance the use of the file. It is more likely to result in considerable frustration on the part of the user"*[5].

As far as we know, the frustration keeps on being the same big problem for current users.

Thus, and considering users and commercial academic editors' trends to Google, and scholar heads' dying efforts to take their users towards subscribed data bases, we propose to focus on three essential steps, within a didactic unit, to ease users step towards efficiently getting pertinent information and contribute to information literacy.

Before going into depth, with the description of the three steps, it is necessary to mention that the present work is based on the published papers in refereed journals that have Impact Factor as a result of a workshop called *Publish Your Research*, taught at the School of Chemistry of the National Autonomous University of Mexico, since 2007.

2. A 3D Self-test Schema
The first step is a self- location test, based on a 3D model concerning the contrast between the academically *disapproved* and *approved* information sources: popular search engines and academic data bases, users´ academic level and the appropriate language users must use. It consists of an electronic

---

Urbach, P.F. Agency Cooperation in Processing Technical Report Literature. 1973. –[http://pubs.acs.org/doi/pdf/10.1021/c160049a008]
[5] Ibid.

application that positions scholar users in their role of retrieving information, weather they tend to use *unaccepted information sources* or prefer to use the traditional subscribed academic database sources.

The self-location test aims to make users aware of their trends towards consulting their library's collections. (See Ibarra, 2010).

The following image shows the sources and their counterparts.

Figure 1.
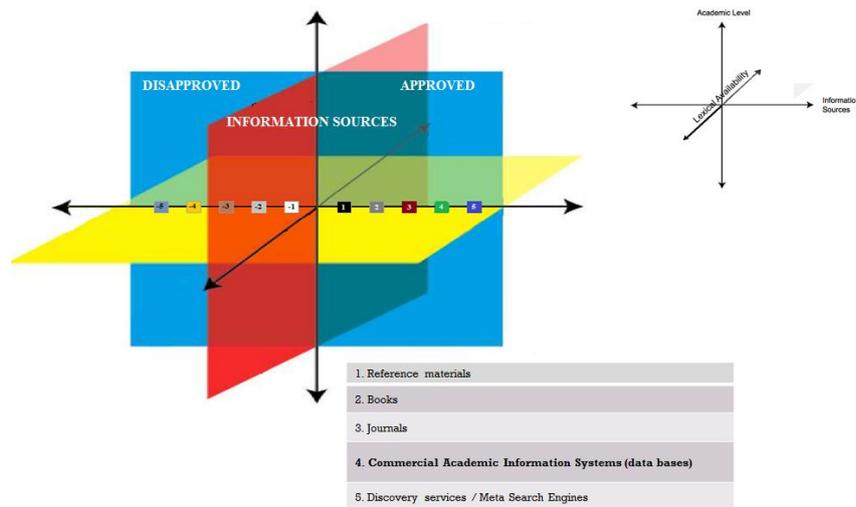
Once users filled in a brief survey, they can see their information retrieval trends. They are asked – and given options - the kind of resources the use, the five described lines above; and the origin of the documents: Library web page or search engines; and, a capital question: where do you take the vocabulary from? Controlled Vocabulary (dictionaries, thesauri, ontologies, MeSH, keywords); Natural Language (false cognates, bad spelling, wrong syntax).

In the following images we illustrate the results of two users in an extreme situation: one who does not use his/her library collections, nor make use of Controlled Vocabulary; and, the other one is quite the opposite.
This action takes place in no more than three minutes.

Figures 2 and 3.
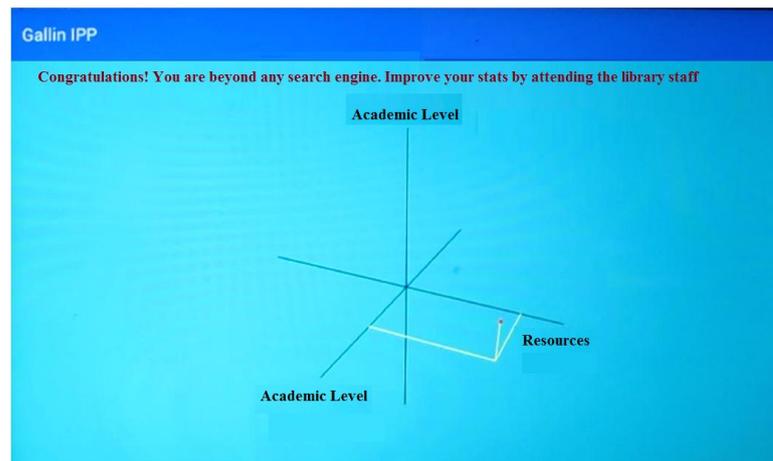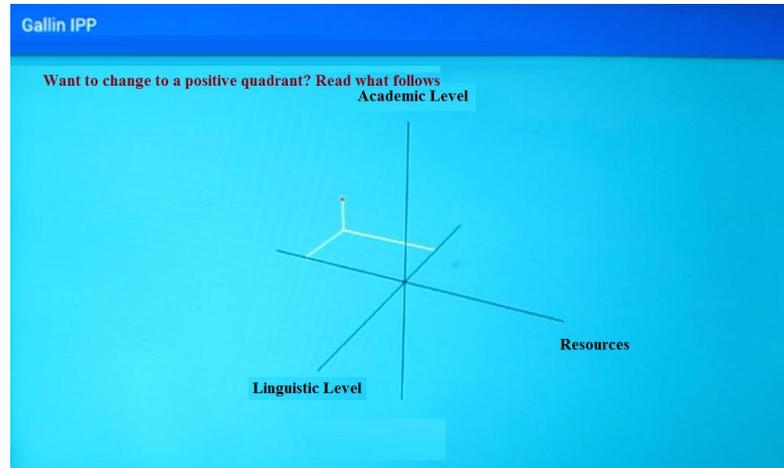
3.   Didactic Procedure to Profit The Library's Collections

Most universities offer their users selective Commercial Academic Information Systems (CAIS), also known as databases, containing substantial information from hundreds or thousands of journals. The great advantages of these resources are their **updated**, **refereed** and **indexed** conditions, which warrantee high quality data. However, in several universities users prefer the "package" that

search engines offer: One single dialogue box, Free resources, Full texts, Quick response, Acceptation of more than one language.

As a consequence, librarians endeavor for publicizing the CAIS's collections in every possible way: posters, workshops, social media, and fora, among some other ways. Teachers and tutors also send their students to the library and warned them not to use search engine documents. Nevertheless, the results are far from being satisfactory.

Considering these aspects, in here it is proposed to approach users with the second step of the present didactic unit: how to present the available databases within the library, focused on users´ FAQs.

Our public university holds more than 150 different databases, which demand to be organized and shown to users in the easiest possible way, that is to say: catalogued. Within the catalogue, users can identify the most appropriate database for their research if they type the TOPIC, name of the database, or the name of the editor. These two last fields are regularly skipped, since users tend to use popular search engines like Google as their information sources. So users may type their topic, and the catalogue will show the related databases, as we can see in the following illustration:

Figure 4.

Thus, by entering the field of interest, users will get the variety their library has to offer in just a pair of seconds. By clicking the numbers on the left, users will see the DB's description in detail. They can see the name, the kind of contents: referential or full text; and, the link.

At this level, we propose to offer a series of exercises related to several fields of knowledge, as well as the answer key.

+ Exercises

> Access the DB catalogue and enter, separately, the following fields of knowledge: Engineering, Biology, Sociology and Literature
>
> Tell how many DB are to each field of knowledge in the catalogue
>
> The answers are shown in the next screen.
>
> **NOTE**: Do not forget to choose the field TOPIC in the corresponding space.

+ Answers

> Engineering   14
> Biology   14
> Sociology   04
> Literature   08

4. Linguistic Steps Towards a Linguistic Storm

The third step for accomplishing is beyond mechanics, and should be brief, in order to accomplish an appropriate timing. It must be focused … *on an educational dialogue with the user in a form of Socratic method where knowledge is that of appropriate expertise, human experience and subjective tool*[6]. *This argument is supported by the fact that* "Koll (1993) found that users provide few clues as to what they want, approaching a search with an attitude of "I'll know it when I see it…" [7]

These two considerations will open the way to develop a research purpose sentence; and, will make it possible to create a linguistic storm, a real and efficient cornerstone to exploit subscribed databases.
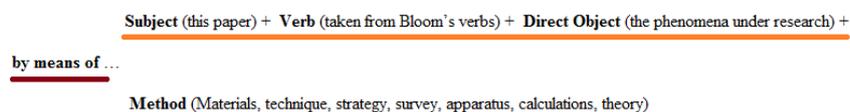
There are certainly problems for English non-native speakers: limited lexicon, lack of thesaurus or some other linguistic tool that ensures a sentence structure

---

[6] Une pédagogie documentaire par le folklore : analyse des modes d'emploi d'Internet au temps de la « frontière électronique »
[7] Koll (1993)

based on their scientific information needs. With this in mind, the sentence must have the following components:

Figure 5.

Subject (this paper) + **Verb** (taken from Bloom's verbs) + **Direct Object** (the phenomena under research) +

by means of ...

Method (Materials, technique, strategy, survey, apparatus, calculations, theory)

For example (Tables 1 & 2):

Table 1.

+ Example of a research sentence

Subject + Verb + Direct Object + by means of + Method

This paper describes
  the isolation of antifungal compounds from the endophyte Nodolusporium sp.
            by means of
the active extract bioassay guided fractionation.

Once in here, users must get the controlled vocabulary terms from a thesaurus, or other linguistic tools.

Keywords
Endophytic fungus
Endophytes
Antifungal
Anti-fungal
Fungicides,
Herbicidal, herbicides
Phytotoxicity

Table 2.

+ Exercise

Subject + Verb + Direct Object + by means of + Method

Subject   :   **This paper**

Verb: _____
Direct Object: _____
by means of
Method _____

Keywords

_____
_____
_____

To fill up the gap of verbs, it is essential to have at hand Bloom's taxonomy, so users would pick up their desired verb. Thus the expression: "*I'll know it when I see it*" will be covered quickly and efficiently.

Table 3.

Knowledge	Count, Define, Describe, Draw, Find, Identify, Label, List, Match, Name, Quote, Recall, Recite, Sequence, Tell, Write 16
Comprehension	Conclude, Demonstrate, Discuss, Explain, Generalize, Identify, Illustrate, Interpret, Paraphrase, Predict, Report, Restate, Review, Summarize, Tell 15
Application	Apply, Change, Choose, Compute, Dramatize, Interview, Prepare, Produce, Role-play, Select, Show, Transfer, Use 13
Analysis	Analyze, Characterize, Classify, Compare, Contrast, Debate, Deduce, Diagram, Differentiate, Discriminate, Distinguish, Examine, Outline, Relate, Research, Separate, 16
Synthesis	Compose, Construct, Create, Design, Develop, Integrate, Invent, Make, Organize, Perform, Plan, Produce, Propose, Rewrite 14
Evaluation	Appraise, Argue, Assess, Choose, Conclude, Critic, Decide, Evaluate, Judge, Justify, Predict, Prioritize, Prove, Rank, Rate, Select, 16

Once the purpose sentence is set, three keywords are selected by users, to create a linguistic storm. Those three words must be entered in a *thesaurus* from the database (or some other linguistic source) to obtain the controlled vocabulary and ensure effectiveness in the information retrieval.

Let us see an example:

Table 4.

Endophytic fungus __A__	*Antifungal* __B__	*Herbicidal* __C__
Endophytes A'	Anti-fungal B'	Herbicides C'
Endophytic fungus [repeated] A''	Fungicides B''	Phytotoxicity C''

As a convention, each keyword is given a letter, and each derived word, is given a ', or '' symbols to indicate the order.

The results of these possible combinations are shown in the next image:

Figure 6.

1st keyword	2nd keyword	3rd keyword	Counts
Endophytic fungus	Antifungal	Herbicidal	29 *
		Herbicides	96
		Phytotoxicity	66
	Anti-fungal	Herbicidal	4
		Herbicides	15
		Phytotoxicity	11
	Fungicides	Herbicidal	14
		Herbicides	111
		Phytotoxicity	63
Endophytes	Antifungal	Herbicidal	22*
		Herbicides	77
		Phytotoxicity	50
	Anti-fungal	Herbicidal	5
		Herbicides	16
		Phytotoxicity	8
	Fungicides	Herbicidal	8
		Herbicides	125
		Phytotoxicity	52
Endophytic fungus	Antifungal	Or Herbicidal	820 *
Endophytes	Antifungal	Or Herbicidal	739 *

*This search led to the discovery of new and interesting papers

The note below the table was written by the young author after using her linguistic storm based on a thesaurus (See: Ibarra (2009)). For the purpose of quickness, we suggest to make the first three combinations: A+B+C (hits); A'+B'+C' (hits) and A''+B''+C''. Please observe that the symbol (+) stands for

the operator "AND", and that the other two operators (OR and NOT) are suggested after doing extra exercises.

What was the result, after two years for this young author, because of the editorial demands, concerning her information retrieval by using a Linguistic Storm?   A Published Paper.

Table 5.

Author:	Rosa Elvira Sánchez-F., et al.
Title	Antifungal Volatile Organic Compounds from the Endophyte
	Nodulisporium sp. Strain GS4d2II1a: a Qualitative Change in the
	Intraspecific and Interspecific Interactions with Pythium aphanidermatum
Article History	Received: 15 June 2015 /Accepted: 15 September 2015
Editor	Springer
Journal	Fungal Microbiology
I.F.	2.973
Key-words	Endophytic fungus . Nodulisporium sp. Hypoxylon anthochroum
	Antifungal .VOCs . Interspecific interaction

This young author expressed her opinion on this linguistic step: *"I don't know why we are not taught this technique before at the university. I'm all of us [Ph. D. students] would save lots of time".*

5. Discussion

Leading users towards exploiting the commercial academic information systems (CAIS) has become a titanic strife on part of librarians, mainly. Now, with three of the most outstanding editors relinquishing their collections to Google and Google Scholar, the efforts to do must be awesome. The challenge has raised expectations; and, the fight between Google and CAIS editors versus Academic Libraries to serve better their users seems to be unfair.

Despite the fact that genius and talent are on both sides, experience is a lot more on libraries´ side, History evidences the scores. Libraries foster a great variety of professional from all fields of knowledge and they can contribute to play an exceptional role in the challenge.

The conditions are clear: users tend to use Googles in an academic environment because they lack familiarity towards library's databases; and, now, editors will impact on decision made to be googable.

Figure 7.
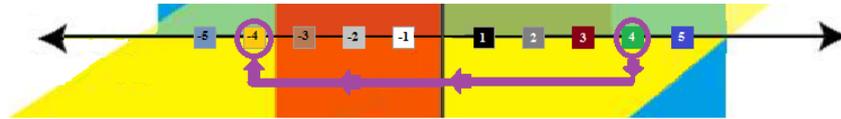
This editors' decision will make their collections, theoretically, widely known, but that does not make users think creatively on their research sentence? Chaos and failure may be present. Some of the three editors' collections are now in both sides, Approved and Disapproved, but the lack on structuring a research sentence is still missing.

From our point of view, some of Googles' Advantages and Disadvantages, in contrast with Libraries' Advantages and disadvantages can be summed up in the following images:

Figure 8.

⁺ Googles

Advantages	Disadvantages	Effects on users
One single dialogue box	Forces users to use "one word" search	Creativity is inhibited
Free resources	Most of them not refereed	Read low quality papers
Full texts	Dubious academic origin	Randomised and weak research
Quick response	Too many unwanted results	Frustration and Waste of time
Accept more than one language	Has not linguistic aids: no thesaurus, no MeSH, no ontologies.	Empoverishes critical thinking and research

Figure 9.

Advantages	Disadvantages	Effects on users
Highly specialised info sources	Force staff to manage them highly specialised	Tend to use Googles
Academic subscribed resources	Poorly publisised	Ignore them
Full & referential texts	Not always at hand	Frustration
Native speakers of local language	Need to expand their liguistic technical horizons	Standby
Staff willing to learn	Need more specialised support on language issues	Become fond of Googles.

6. Conclusions

Using mobile gadgets is a most in current times for libraries, but most of all, some didactic approaches are definitively essential. The app presented in here is certainly very basic and must be adapted into the libraries perspectives, languages and collections, but that is not enough.

Users awareness of numerous information resources is an acceptable beginning; on the other side, reference materials, books, journals, especialised data bases and discovery services can be consulted through some clicks, only if users know the names, the titles and specific topic, but there exists an imperative sentence structure that is missing.

Finally, two additional advantages derived from this brief study are: the research sentence and linguistic storm can be used in any information source, approved or disapproved; and, the data collected by the app can monitor users literacy, development and advances towards their library's collection.

References

Bloom's verbs. Webb's Depth of Knowledge. http://plaeadok.weebly.com/uploads/1/3/1/9/13192411/blooms_verbs.png

Dirección General de Bibliotecas web site: [http://www.dgbiblio.unam.mx/index.php]

Does Google Scholar link with EBSCO? (2016) http://support.ebsco.com/knowledge_base/detail.php?id=3590 Accessed March, 8, 2016.

Ibarra, R. (2009). Linguistic storm. Linguistic storm: an essential information retrieval tool to update researchers (pp 97-106). *QQML- e-journal*. http://www.qqml.net/index.html

Koll (1993), en Griffiths, Jill R. and Brophy, P. "Student searching behaviour in the JISC Information Environment". 10-Oct-2002. *Ariadne Issue 33*. URL: http://www.ariadne.ac.uk/issue33/edner/

Paloque-Berges Camille (2013) «Une pédagogie documentaire par le folklore: analyse des modes d'emploi d'Internet au temps de la «frontière électronique». *Documentaliste- Sciences de l' Information* 4/2013 (Vol. 50), p. 64-71 www.cairn.info/revue-documentaliste-sciences-de-l-information-2013-4-page-64.htm. DOI: 10.3917/docsi.504.0064.

Science Direct-ly into Google (2007). [http://toc.oreilly.com/2007/07/science-directly-into-google.html] Accessed March, 8, 2016.

Urbach, P.F. (1973) Agency Cooperation in Processing Technical Report Literature. *Journal of Chemical Documentation. Vol. 13. No. 2, 1973. P. 69-72.* http://pubs.acs.org/doi/pdf/10.1021/c160049a008

Why ProQuest is working with Google to improve research workflows (2015). http://www.proquest.com/blog/pqblog/2015/Why-ProQuest-is-working-with-Google.html   Accessed March, 8, 2016.