Qualitative and Quantitative Methods in Libraries (QQML) 4: 97-– 106, 2015

Linguistic storm: an essential information retrieval tool to update researchers

Rafael Ibarra Contreras
Dirección General de Bibliotecas, UNAM.

Abstract: Shows linguistic storm's utility and efficiency through controlled vocabulary matrices. A controlled vocabulary matrix construction is described, as well as the obtained results from different academic and commercial information systems. The obtained results indicate the semantic universe that researcher must consider to confirm, modify or change their research papers and trends.
Keywords: Information Retrieval, IR, Controlled vocabulary, Natural language Information Systems, Linguistics, Lingstorm, Semantics, Infopragmatics, Pragmatics

1. Introduction

Young researchers wish to be updated on their field of knowledge at an international level to warrant the handling of the emerging science trends, to publish their findings in indexed traditional or open access journals, to high-rank their schools as a result of publishing productivity and to assure their school grants. However, to achieve these goals, young researchers must face three barriers: chaotic interfaces of the Commercial Academic Information Systems (CAIS) and severe limitations of linguistic help; English language appropriate management and absence of controlled vocabulary awareness. In this paper we suggest a linguistic storm (*Lingstorm*) as an efficient solution to beat these three hurdles.

CAIS' chaotic interfaces present dozens of elements that are of poor utility to novice users and "*a cumbersome key (F1) that clearly shows how unsuitable they might be for the numerous users kinds"*, Ibarra, (2010). Furthermore, as Katsirikou and Skiadas (2001) point out, there are 23 processing actions that comprise the opening and the closing dialog in an information request that go from finding the appropriate electronic resource to indirectly and unwittingly provide personal information on one's activities, no matter the language. Expert users may as well avoid those elements that are of no interest for them, but what can novice users do?"

Received: 21.5.2014 / Accepted: 1.12.2014 ISSN 2241-1925

On the other hand, as science is internationally developed by different researchers, and into different languages and characters, young researchers (YRs), whose native language is not English, commit a big variety of mistakes: bad spelling, wrong affixation, iterative use of natural language, and the absence of lexical availability to establish and plan their search strategies. So, with no plan and lack of linguistic availability in English, what YRs may use as search words?

Generally, controlled vocabulary is acquired along the years of reading specialised literature and attending school and fora related to their academic fields. As undergraduates, masters or Ph. Doctors, YRs may spend several years to achieve a satisfactory linguistic competence.

When YRs start to search for information, they usually overlook controlled vocabulary terms, mainly because of two reasons: 1) they do not know linguistic tools (thesauri, ontologies, subject headings, indices, specialized dictionaries and key words). 2) They simply "feel lucky" and begin two common unwritten basic strategies: top-down, they restrict their search by collecting "new and appropriate" terms as they appear while they read; or bottom-up, doing the reversal method, from a specific term to a more general one. In my experience, as an instructor in a workshop called *Publish Your Research in Indexed Journals*, when approaching to the linguistic tools, YRs admit they are not familiar with them; they also admit to utilise no method or technique for their information search. This last remark is quite similar to the one Griffiths and Brophy (2005) point out in their report: *"Some admitted that they simply did not have any further search strategies, saying they "Don't know where else to search for it," "I have searched everywhere I can think of," or "didn't know where else to go."*

However, with the linguistic tools describe above, YRs may break the rule of spending several years before becoming familiar and manage controlled vocabulary by means of a tool called linguistic storm (*Lingstorm*).

2. A Linguistic Storm

A Lingstorm is a matrix composed of a combination of controlled vocabulary terms derived from the user's research purpose. That is, YRs must write a **verb** that best describes their purpose and the direct object of the verb, followed be the phrase *by means of* and complemented by their method, instrument or technique they used. In the following example[1], **CASE 1**, we can see:

To assess metallurgic slag as Fenton-like catalyst on disinfection water by means of slag- H_2O_2- solar light

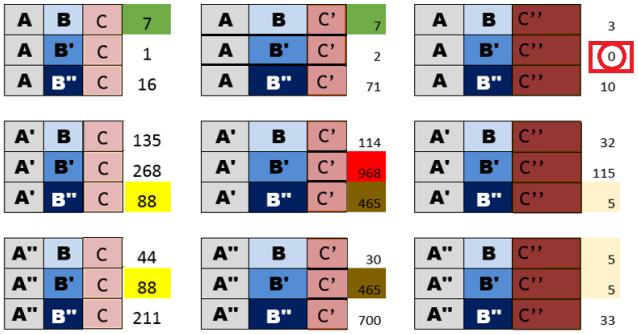
From this sentence, YRs must take those terms they consider as having the key

[1] This was a real example taken from a young researcher enrolled in the workshop, Publish Your Research in Indexed Journals, offered by the School of Chemistry. At the appendix, there is a brief survey.

weight; and, then, have some of the available linguistic tool to obtain a controlled vocabulary matrix.
For example:

Fenton A	**Disinfection B**	**Photocatalysis C**
Oxidation A’	Cleaning B’	Catalysis C’
Catalytic oxidation A’’	Water treatment B’’	Photochemical reactions C’’

Then, YR must do the 27 combinations of the elements involved. The terms, from A’ to C’’, were collected from the thesaurus integrated in the Engineering Village Database, from Elsevier. The hits were as shown in the following table:



The YR comments expressed on the use of the matrix were the following: The CAIS used was Engineering Village 2, by Elsevier.
The search was useful for my paper. If I use only two terms the results were

- Fenton and disinfection: 30 papers
- Fenton and photocatalysis: 21 papers
- Disinfection and photocatalysis: 21 papers

By combining the main terms with the related terms, I obtained:

- **Fenton** and **cleaning** and **catalysis**: 2 papers
- **Fenton** and **water treatment** and **photochemical reactions**: 9 papers
- **Oxidation** and **photocatalysis** and **photocatalysis**: 166 papers.

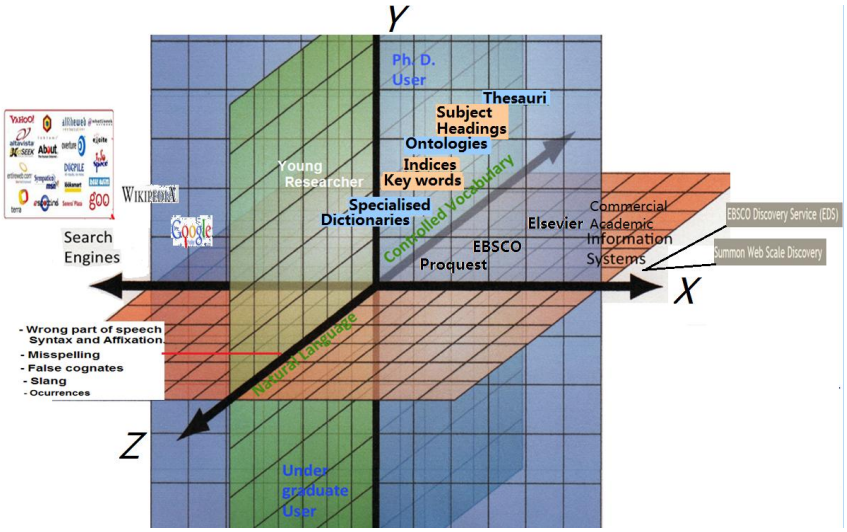
By using the thesaurus, I could refine the papers search related with my field of study. I found, at least, two papers of great interest. I could identified a candidate journal to publish my research: Chemosphere.

Considering the previous experience, we are in a position to reflect on the involved actors’ results by means of 3-dimension model self-test that can be

used as a self-test, regarding the information retrieval habits anyone may have and in almost any language.

3. Information Retrieval Self-Test

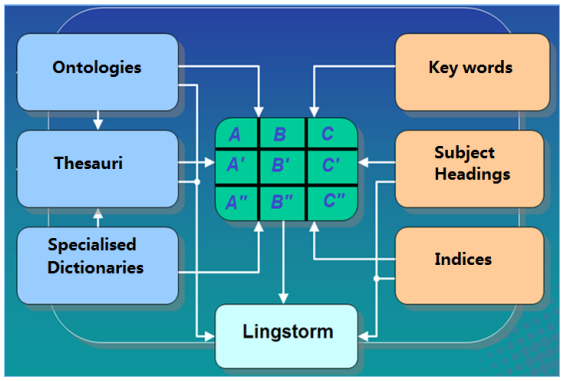
In almost every search strategy, we can easily identify the most common elements that can influence, for better or for worse, the experience in IR. These elements are, on axis **X**, from the origin to the left, the search engines that are usually banned from tutors to YRs to retrieve information. On the opposite side, the CAIS that are accepted, including the emerging Discovery Services systems. On axis *Z*, from the origin to the front, the use of natural language is described; in the opposite direction, the controlled vocabulary and the tools that can ease lexical availability to build up a *Lingstorm*. In addition, on axis *Y*, the humans involved on satisfying their information needs, according to their academic level.



With this IR3DM[2], almost all YRs can interestingly self-test their ability to retrieve pertinent information.

So far, by creating a *Lingstorm* the results of some YRs (Spanish Native Speakers) attending a publishing workshop have reported interesting clues that can be of great help to their colleagues whose native language is not, but may be, English.

[2] The 3d image, not the texts, is credited to Sakurambo, http://commons.wikimedia.org/wiki/File:3D_Cartesian_coordinates.PNG#mediaviewer/File:3D_coordinate_system.svg



Among some other thought-provoking reports, here is a brief description of some *Lingstorms* results and comments (recorded in Spanish) from the YR involved. The following one is related to a research done on Psychology:

CASE 2

"The results may show countless records and that my Lingstorm was useless, but let me tell you that as I am new to the field, I was curious to know how much data could be found by using general terms. So, instead of feeling disheartened, I felt happy because now I know more about my field of interest!

The purpose was:
To characterize the effect of 5-HT prenatal deprivation during different rat gestational stages by means of tryptophan free diet on the adult pyramidal neurons morphology.

The corresponding *Lingstorm*:

A Serotonin	B Cerebral Cortex	C Development	27929
A´ 5-HT	B´ Brain Cortex	C´ Maturation	4901
A´´ Monoamines	B´´ Cortical	C´´ Ontogeny	1304

A	B	C	27929
A	B	C´	6417
A	B	C´´	3642

A´	B	C	27929
A´	B	C´	3612
A´	B	C´´	2094

A´´	B	C	7497
A´´	B	C´	1805
A´´	B	C´´	1347

A	B´	C	42427
A	´B´	C´	8760
A	´B´	C´´	5044

A´	B´	C	23607
A´	B´	C´	4901
A´	B´	C´´	2930

A´´	B´	C	2270
A´´	B´	C´	2270
A´´	B´	C´´	1731

A	B´´	C	32371
A	B´´	C´	7304
A	B´´	C´´	3849

A´	B´´	C	19490
A´	B´´	C´	4497
A´	B´´	C´´	2389

A´´	B´´	C	7663
A´´	B´´	C´	1774
A´´	B´´	C´´	1304

The findings:
I became familiar with Elsevier, ScienceDirect and Ovid. I already knew PsycINFO. As well, I was able to know some journals and their corresponding IF:

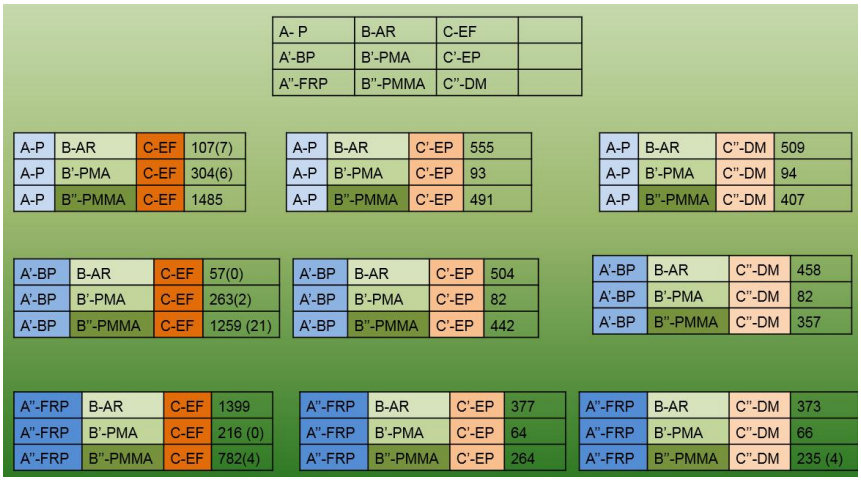
Journal	Impact Factor
Cerebral Cortex	6.828
Frontiers in Cellular Neuroscience	4.5
European Journal of Neuroscience	4.345
Developmental Brain Research	2.892
International Journal of Developmental Neuroscience	2.692

The last journal is the one my tutor and I are interested in sending our research paper.

CASE 3
I used Engineered materials CAIS. The purpose of this paper is the polymerization of methyl methacrylate monomers by means of electric field. The Lingstorm was an interesting exercise.

Thesaurus selected terms:

Polymerization (P); Bulk polymerization (BP);
Free radical polymerization (FRP); Acrylic resins (AR)
Polymethacrylates (PMA); Polymethyl methacrylates, (PMMA);
Electric Field (EF); Electric polarization (EP)
Dipole moment (DM).



Findings: *Next to each series of hits, I noted the probable abstracts to be considered in my paper.*

CASE 4

The purpose of this paper is to describe child human rights abuses identified by means of medical records review by trained physicians.

a minor	b Human rights	c abuses	d physician
a' child	b' rights	c' violations	d' paediatricians
0	b" Convention on the rights of child	0	0

a	b	c	d	4 (1)
a'	b'	c'	d'	0
0	b"	0	0	
a	b'	c	d	4 (repeated)
a	b"	c	d	1
a	b	c'	d	2 (0)
a	b'	c'	d	As above
a	b"	c'	d	0
a	b	c´	d´	0
a	b´	c´	d´	0
a	b´´	c´	d´	0
a'	b	c	d	21(5)
a'	b'	c	d	0
a'	b"	c	d	0
a'	b	c'	d	9 (1)
a'	b'	c'	d	9 (exactly as above)
a'	b´´	c'	d	0
a'	b	c'	d´	1
a'	b´´	c'	d´	0

This YR reported 21 comments and eight conclusion points. As a brief resume, I included the most illustrative. Findings and comments:

The term B´´ (Convention on the rights of child) is not a MeSH term, but it used by the UNO to refer the rights of the child. The term is used since November 20, 1989. On the combination a'+b'+c'+d'=0, I put out the term physician (d') and I found a very good paper. This may support the discussion and the project. In this combination, I found the most specific papers; besides, what I expected to find in my search. The article by Weintraub deals with the efficiency within the medical-legal relationship in childcare to improve children health.

To do all this task took 7.5 h, including this writing report for the workshop, reading some available abstracts and the request of several papers to the library. In a previous exercise, by using a partial Lingstorm, it took 8 h. just to have a general look (not reading or browsing the papers). I did not get expected results.

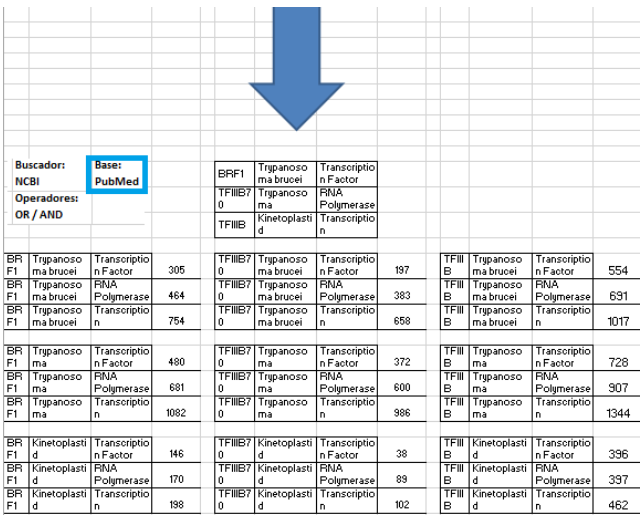
CASE 5

This last case was on microbiology. Purpose: *To characterize the BRF1 ortholog in Trypanosoma brucei by means of defining three transcription factor-specific characteristics, i.e. nuclear localization, association with BDP1 and TBP, and participation in RNAP-III transcription.*

Buscador:	
NCBI	**Base: PubMed**
Operadores:	
AND / AND	

BRF1	Trypanosoma brucei	Transcription Factor	
TFIIIB70	Trypanosoma	RNA Polymerase	
TFIIIB	Kinetoplastid	Transcription	

BRF1	Trypanosoma brucei	Transcription Factor	0	TFIIIB70	Trypanosoma brucei	Transcription Factor	1	TFIIIB	Trypanosoma brucei	Transcription Factor	1
BRF1	Trypanosoma brucei	RNA Polymerase	0	TFIIIB70	Trypanosoma brucei	RNA Polymerase	0	TFIIIB	Trypanosoma brucei	RNA Polymerase	0
BRF1	Trypanosoma brucei	Transcription	0	TFIIIB70	Trypanosoma brucei	Transcription	0	TFIIIB	Trypanosoma brucei	Transcription	0
BRF1	Trypanosoma	Transcription Factor	0	TFIIIB70	Trypanosoma	Transcription Factor	0	TFIIIB	Trypanosoma	Transcription Factor	1
BRF1	Trypanosoma	RNA Polymerase	0	TFIIIB70	Trypanosoma	RNA Polymerase	0	TFIIIB	Trypanosoma	RNA Polymerase	1
BRF1	Trypanosoma	Transcription	0	TFIIIB70	Trypanosoma	Transcription	0	TFIIIB	Trypanosoma	Transcription	2
BRF1	Kinetoplastid	Transcription Factor	0	TFIIIB70	Kinetoplastid	Transcription Factor	0	TFIIIB	Kinetoplastid	Transcription Factor	0
BRF1	Kinetoplastid	RNA Polymerase	0	TFIIIB70	Kinetoplastid	RNA Polymerase	0	TFIIIB	Kinetoplastid	RNA Polymerase	0
BRF1	Kinetoplastid	Transcription	0	TFIIIB70	Kinetoplastid	Transcription	0	TFIIIB	Kinetoplastid	Transcription	0



The YR findings and comments refer that in his first *Lingstorm* he just used key words.

I did two Lingstorms, both of them using PubMed, but in the first search the Boolean operators were AND, AND to join the terms; in the second, AND and OR were used. I did not use any thesauri. The used terms were taken from different papers.

The search that provided many hits are due to the general used terms. So to restrict the search I used AND, AND. I discovered that the six hits were related to papers my colleagues and I already knew. This issue made us feel relaxed because there were "no new news". I say this because we know that there are two American groups in competence to publish their findings; and, we want to participate in that competence.

4. Conclusions

The previous YRs' examples and comments show, in some way, how they can be aware of the IR scenario and take the necessary steps to prevent frustration and waste of time by using an *Information Retrieval Self-Test*, and build up a *Lingstorm* as an efficient linguistic tool. The series of hard data obtained from CAIS allowed YRs to confirm their pursue of offering a valuable paper to their selected editors; to modify an incomplete or obsolete research; and to change the focus of their scientific aims. Among the benefits, we might remark that YRs may get earlier to research maturity; *Lingstorm* allows master trends, processes of international research, lead YRs to be in the fittest environment for their purposes, reinvigorate the research duties as stimulating ability rather than a heavy task that is possible if they follow stiff linguistic articulation can do. After all, knowledge is not originated in a specific language. *Lingstorm* gives a useful meaning to IR.

References

Griffiths Jillian R. and Brophy Peter, (2005). Student Searching Behavior and the Web: Use of Academic Resources and Google. Library Trends, Vol. 53, No. 4, 539-554.

Ibarra, R. (2010). *Infopragmatics: an efficient method for information retrieval: Linguistic solutions for pertinent information*. Lambert Academic Publishing. Germany.

Katsirikou, A. and Skiadas C. H., (2001). Chaos in the library environment, *Library Management*, Vol.22, No. 6/7, 278 – 287.

APPENDIX

Survey answered by five YRs enrolled at the workshop Publish Your Research in Indexed Journals:

QUESTION	ANSWER	RESULTS
1. How did you like the *Lingstorm* use for **pertinent** Information Retrieval?	Totally useful_____ ☑☑☑☑☑	5
2. Did you use an Information Retrieval tools before using a *Lingstorm*?	No _☑☑_☑_☑_☑ Yes____ (Which one?_____)	5
3. Did you know the Thesaurus as an implement to enhance your information search?	No _☑_☑_☑ ☑ ☑ Yes____ (Which one?_____)	5
4. *Lingstorm* helped you to … You can select more than one case.	Change ___☑_☑ Confirm_☑_☑_ Modify _☑☑_☑ ☑ ☑ your research paper.	2 2 5
5. How long did it take you to build up your controlled vocabulary matrix?	2 h. 2 h. 8 h. 5 h. 2-4 h.	4 hours average.
6. How long did it take you to complete your Information Retrieval based on the controlled vocabulary matrix?	8 h. 7 h. 8 h. 8 h. 6 h.	7.4 hours average.
7. Did you get acquainted with any new journal because of using your *Lingstorm*?	1_☑ 2____ More than 2__☑_☑_☑ ☑	1 0 4
8. Did you find an article to be improve your research because of using a *Lingstorm*?	1_☑__ 2____ More than 2 _☑☑_☑_☑_	1 0 4