

Learning about Sancho Panza and foreign cousins : a case for introducing the future of e learning within the semantic web

Introduction

Many e learning project researchers and managers agree upon the need of having the learners work together as a community of learning. “Co-operation”, “community”, as well as “communicate” has to do with “having something in **common**”. Within a multi-origins group of learners and lecturers, people have a reduced version of the working language in common. But is language itself, as a set of words, the only way for understanding each others? I’ll propose to consider what learners from the Large Europe and Mediterranean area may have in common “before”, even if they are not aware of.

Then I shall consider the process in which they construct a common set of reference texts through browsing the web. The actual means and tools for retrieving are compared with the larger capabilities of the semantic web.

Before Babel : what the Euro Mediterranean learners may have in common

My experience is that, whatever the subject of the course may be, learners arrive at a point where they ask the question “what have we got in common?” They quickly discover that most of them use Arab words like Algebra, Indo-European words like Guitar, and Greek words as well and that, behind these words there are :

- Common conceptions that the elements of the world can be measured (Algebra, etc.).
- Common patterns for exchanging through music and words (the guitar allowing to play Arabo Andalus music as well as Rock and Roll).
- Common references to the writings of Plato and Aristotle directly or through local further writers.

Eventually, one participant may underline that they have something else in common, knowing that year 1996 has been proclaimed Nasreddin Hodja Year by UNESCO. Nasreddin is one of the names of a character within stories that are universal because they describe human nature and weaknesses of mankind everywhere. The second main orthography of the name is Nasrudin. This universal character inspired Cervantes for the actions and thinking of Sancho Panza. Their respective popularity can be verified on the Internet, with Google search engine for example.

The number of documents for Nasreddin is about 10 thousands, for Nasrudin and Nasruddin about 5 thousands and 200 thousands for Sancho Panza!

How amazed are the learners! At first glance, they thought they were so different. Speaking different languages, having different folkloric costumes, having different religions, having different political systems in their home countries. Having Algebra, Minor scales and Philosophical concepts in common is somehow “sophisticated”. But they realise that they laugh at the same character about the same situations!

Typical of this character is that he has a donkey. Let me tell a true story that happened in a learning community. A fellow, within the learning community, tells a story about Nasreddin, the donkey and the custom keeper. And a guy from Belgium, opening wide eyes, says to the group : “In Walloony, we have this same story but the hero is a Walloon and the donkey is replaced by a bicycle!” (1). Anthropologists observe that this common set of stories span from Ireland to North India. Other researchers investigate on what all human beings have in common that is not visible at first glance. Searches on the “original language” face methodology difficulties –Abehsera.

More promising are the works on the common construct of languages with metaphors –Lakoff & Johnson (1980). The work on ancient fairy tales –Propp (1927) and followers is well known. The facts explored by Jaynes are fascinating as well as those by Girard.

We shall keep Nasreddin, with the great number of stories he appears in, as our common core of texts retrieval example.

Retrieving for exchanging

In a community of learning, one major activity is to retrieve scientific information. Each learner has “ideas” and needs to enlarge, improve, strengthen his/her views, it is what learning is about. Before Guttenberg, the source of knowledge was people and hand written documents, then printed books and reviews filled libraries. In the 70’s, information networks began to store abstracts then full text. The first generation of the Internet is about full text and dynamic links between texts.

Retrieving from Dialog : the “old” way

Among different similar services, Dialog proposes “12 terabytes of content from the world's most authoritative publishers, and the products and tools to search every bit of it with speed and precision” (2).

Before the spreading of content through the Internet, Dialog was the main way to retrieve a set of texts on a selected matter. Most databases within the Dialog ensemble are made of text with a keywords section for each. These keywords were/are given by the authors and/or reviewers following strict rules of the place of keywords within a thesaurus, i.e. a tree organising classes and sub classes of objects or concepts. (3)

Retrieving with Google : finding popular documents, missing pertinent documents

In 2002, how is it possible to find the pertinent documents on Nasreddin?

I choose to talk about *Google* because it has the highest number of trophies in the high speed search engine category (4) (5). Up to the need of the retriever, other tools may be useful :

- Other search engines (All the web, Altavista, Excite, Inktomi/Hotbot, Infoseek,).
- Engines that use the content of other engines (Metacrawler, Ixquick).
- Directories (Lycos).
- Human-compiled directory of web sites (Yahoo, LookSmart/MSN).
- Human-powered search service (Ask Jeeve).
- Volunteer editors catalogue (Open Directory/Netscape).

Searching for “Nasreddin”, we already saw that *Google* finds 10 thousands links. *All the web* proposes 6 thousands. In any case, the volume of the finding is the problem, a two fold problem. First issue is to focus the research i.e. to eliminate all the answers that don't fit with our issue which is : “What are the interesting texts about this cultural set of stories/views of the world that we have in common”. Among the links that we want to eliminate are, for example :

- the restaurants that are called “Nasreddin” (160+).
- the “Sancho Panza” cigars adds (more than 7 thousands out of the 10!!!).
- the documents about “Nasrudin” and “Buddhism” (219).

The second issue is the ranking of the documents. The choice of Google managers, for example, is to rank the documents through popularity i.e. the number of other documents that give links to the document. And the popularity of these pointing documents is itself evaluated.

So the popularity given by Google has pretty good reliability.

But is this ranking adequate for the learner or is the pertinent document hidden within the list?

Shall we consider the documents selected by Google for keyword Nasreddin.

The first and eleventh documents give the description of the historical Nasreddin and stories (6). This is an interesting beginning. For Sancho Panza, it is a lot more difficult. The information is lost in a list with documents about cigars, sculptures, restaurants, other authors speaking of Sancho (ex. Kafka), music, etc..

If we focus our research, using a formula with three scope reducers :

- documents containing "sancho panza character".
- excluding documents with “-cigar -corona -bachilleres”.
- keeping documents written in English only.

Then, only three documents remain, two about music and one text for students. It's pretty poor. If we focus “Sancho Panza origin”, “Sancho Panza filiations”, “Sancho Panza history”, no document is found. Here is an illustration of the retrieving problem within the actual Internet. Researchers still need to use Dialog databases because of the capacities for searching within different areas (key words, full text, etc.).

On Internet, a formula with Nasreddin and Sancho Panza points to a very interesting non academic document on the universal character of the “holy fool”. But we found it partly because, through Dialog database research, we new the answer, i.e. that Nasreddin and Sancho are of the same nature. Some researchers talk about Naredin or Sancho when studying “court jesters”. (7)

The future Internet should combine the possibilities of the actual Google, of directories and Dialog.

The Holly Grail : the semantic web with ontologies

Both humour and wisdom of Cervantes story comes from the gap between the Holly Grail that Quixote searches and Sancho's Terrestrial Grail. In his peak moment of inspiration Sancho says to Quixote his dream of being the governor of an island. What are the dreams of curious learners as well as publishing scholars? They are starving for the tool that will allow them to find the right rare document quickly and accurately.

The art of tagging

Within the next ten years, one of the big deal of the Internet will be to organise the tagging of documents. Tags, as their name suggest, are hidden labels that are put “behind” the text of a document. For example, when having an HTML file on the screen with a browser or a text processor it's generally possible to ask the view of the HTML code. In the middle of this code one can recognize the main text of the document. At the beginning, there may exist tags like <title>, <author>, <last author>, <revision>, <created> or <last saved> followed by the corresponding

information. The evolution of the web will be that these tags will be more and more used by software applications and that more and more types of tags will be available.

For example, within collaborative learning, collaborative writing is an interesting activity. The availability of the tags previously described allows a software application to manage the different versions of a document and the follow up of the intervening authors. (8)

For the retrieving issue, tags allow many possibilities.

Today (2002) if one searches “Berners Lee” –the name of the inventor of the HTML language in the 80’s and actual boss of the World Wide Web Consortium- about 100 thousands of documents are proposed. When the tagging systems that already exist (XML language complementing HTML) will be used, it will be possible to select only the documents authored by Tim Berners Lee. It will be a tremendous progress.

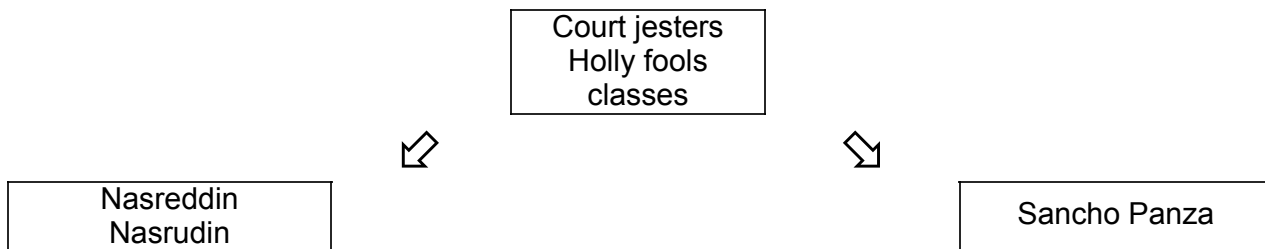
If we consider the future relationship between authors and learners, we can imagine a future of dynamic matching. For example, shall we imagine Tim Berners Lee writing about the semantic web. He would prepare paragraphs, drawings, spreadsheets and organise them with a set of rules. When a learners searches “semantic web” he is asked to be more precise about his concern, or he has his interest profile permanently defined. What he will receive is a customized article made of the pertinent items (paragraphs, drawings, etc.). BOIS (2002).

These are only small examples of what tags allow to do both for authoring and retrieving. (9)

The art of meaningful tags

The tags we have seen before are only information tags, they don’t carry any meaning. An author’s name, a document characteristic (date, version, language, medium, etc.) doesn’t give more than the label’s content.

Shall we come back to our core example.



This diagram is a part of an ontology.

“An ontology defines the terms used to describe and represent an area of knowledge. Ontologies are used by people, databases, and applications that need to share domain information (a domain is just a specific subject area or area of knowledge, like medicine, tool manufacturing, real estate, automobile repair, financial management, etc.). Ontologies include computer-usable definitions of basic concepts in the domain and the relationships among them.” (10)

If we imagine the documents about Nasreddin and Sancho within a few year, these document will have a tag like <ontology> = the address of a document within an academic site specialized in this matter

<place1>= subclass of “Court jesters”

<place2>= subclass of “Holy fools”

The future browsers will have a button that will automatically call for the ontology and allow the learner to decide if the domain where the search is automatically oriented fits with his/her needs.

The effort for the Holy Grail

As described before, the semantic web, i.e. with ontologies is very simple to implement.

First, learned societies develop domain ontologies i.e. lists of allowed terms, their relations and definitions. Second, authors, using the new tag editing software applications, complete their text with tags. Third the retrievers use the new browsers that allow to select documents by specifying tag contents and relations. All this is simple, it doesn’t mean that there is no effort. There is a quantitative effort for producing the ontologies and tagged documents. Qualitatively, there are many cases where consensus should be found between searchers within a same learned society and between learned societies to have one element described with the same name in physics and chemistry, for example.

Automated thinking

I have developed, here, only the first step of the semantic web. More important is the ability of computers to manage documents or sub-documents with contents and meaning defined by tags referring to ontologies. Both with ontologies –which carry the meaning- there is the need for a coherent syntax. Different bodies, including the Dublin Core Metadata Initiative (11), work on this matter.

Conclusion

One of the core need for e learners is to work within communities of learning which share knowledge i.e., primarily, documents. This issue is even more important in the case of multicultural groups. The actual and future main source of documents is the Internet. Actual bad accessibility of documents on the web is a temporary problem caused by low reactivity of actors toward this problem that was known long before when documents had to be retrieved in large libraries or institutions computer storage. The age of syntax and semantic standardization for the retrieval process on the Internet has begun and this problem is to be solved at the machine level. Then, all scientific writers and learners will have to learn how to use these new facilities.

Notes :

(1) The story can be found in a book by Paul Watzlawick, "The language of change".

(2) <http://www.dialog.com/>

(3) Database producers as IEE for INSPEC database or MedlineUSA for MEDLINE have staff of one to several hundreds of people dedicated to indexing. Dialog includes hundreds of databases. Yahoo has more than one hundred people for indexing, etc.. How many people in the world index documents every day?

(4) Studies about search engines performances can be found at

<http://searchenginewatch.com/reports/index.html>

The history of the creation of Google within Standford University can be found at :

<http://www-db.stanford.edu/pub/voy/museum/google.htm>

Google includes a mechanism similar to the one of directories which allows to select within an area of interest :

Arts

[Movies](#), [Music](#), [Television](#), ...

Business

[Industries](#), [Finance](#), [Jobs](#), ...

Computers

[Internet](#), [Hardware](#), [Software](#), ...

Games

[Board](#), [Roleplaying](#), [Video](#), ...

Health

[Alternative](#), [Fitness](#), [Medicine](#), ...

World

[Deutsch](#), [Español](#), [Français](#), [Italiano](#), [Japanese](#), [Korean](#), [Nederlands](#), [Polska](#), [Svenska](#), ...

Home

[Consumers](#), [Homeowners](#), [Family](#), ...

Kids and Teens

[Computers](#), [Entertainment](#), [School](#), ...

News

[Media](#), [Newspapers](#), [Current Events](#), ...

Recreation

[Food](#), [Outdoors](#), [Travel](#), ...

Reference

[Education](#), [Libraries](#), [Maps](#), ...

Regional

[Asia](#), [Europe](#), [North America](#), ...

Science

[Biology](#), [Psychology](#), [Physics](#), ...

Shopping

[Autos](#), [Clothing](#), [Gifts](#), ...

Society

[Issues](#), [People](#), [Religion](#), ...

Sports

[Basketball](#), [Football](#), [Soccer](#), ...

(5) I made a test on Nasreddin within the Open Directory <http://dmoz.org/about.html>

The results were not pertinent.

(6) It is interesting to compare two versions of the same "knowledge."

First version at : http://w1.871.telia.com/~u87109316/index_eng.htm

Nasreddin Hodja was a cleric during Seljuk times. He was born in 1208 in Hortu village near Sivrihisar in Central Anatolia. As a young boy he must have enjoyed a free country childhood and lived in one of the cottages with adobe walls and flat baked earth roofs, typical of this region. He received his early education from his father, the village imam, and went on to study at the medrese.

After working as a village imam for some years, he moved in 1237 to the town of Aksehir. There he is known to have studied under such notable scholars of the time as seyid Mahmud Hayrani and Seyid Haci Ibrahim. Later he became a professor at the medrese in Aksehir and served as kadi. Nasreddin Hodja died in 1284 at the age of 76, and was buried in Aksehir in a tomb which symbolizes the absurdity in life which he had loved to expose while alive. A door with a great lock stands by the tomb, but there are no walls for a door.

Second version at : <http://www.turkey.org/groupd/chapter1/nhodja.htm>

The oldest Nasreddin Hodja story is found in the book called "Saltukname" written in 1480, which also contains other folk stories and legends. It is stated in "Saltukname" that Hodja was born in Sivrihisar and that the natives of Sivrihisar were famous for their strange behavior and ingenousness. The strange behavior of the natives of Sivrihisar is also mentioned in a handwritten story book in Biblioteque Nationale in Paris. These documents are considered proof of his birth in Sivrihisar.

(7) <http://www.omphalos.net/files/shaman/FOOL.TXT>

(8) Collaborative software links

<http://wwwfac.worcester.edu/owl/tech/techwrite.htm>

(9) Tim Berners Lee about publishing in the semantic web era

<http://www.nature.com/nature/debates/e-access/Articles/bernerslee.htm>

(10) Defining ontology

<http://km.aifb.uni-karlsruhe.de/owl/>

Example of the chemical markup language

<http://www.xml-cml.org/schema/CML2/Core> in <http://www.ch.ic.ac.uk/rzepa/codata/>

Scientific publications in XML - towards a global knowledge base.

Peter Murray-Rust,^a Henry S. Rzepa^b

(11) <http://dublincore.org/documents/dcmes-qualifiers/>

References :

Abehsera Abraham *Babel Language of the 21st Century* edited by A. Sutton (Ekev).

B0IS Christian (2002) *Who Should "Customise" the Knowledge Content? Publishing Scholars or On Line Educators?* Eden 2nd workshop March 2002 Hildesheim

Girard, René *Violence and the sacred* Patrick Gregory (Translator) (Paperback - February 1979)

Jaynes Julian (1976) *The Origin of Consciousness in the Breakdown of the Bicameral Mind* Houghton Mifflin Company Boston ISBN 0 395 56352

Lakoff George, Johnson, Mark (1980) *Metaphors We Live By* Univ. of Chicago Press (Chicago) Summary in <http://endeavor.med.nyu.edu/lit-med/lit-med-db/webdocs/webdescrips/lakoff1064-des-.html>

Otto, Beatrice K. *Fools are everywhere* History Today, June 2001

Propp Vladimir (1927) *Morphology of the folktale* Austin, TX: University of Texas Press, 1968