The Precision of Metaphor for Information Retrieval

Abstract
We present a theoretical model for an information retrieval filter based on metaphors. The justification for this proposal is the current information boom, especially due to new technologies. The theoretical basis for our proposal is an association of the Cooperation Principle proposed by Grice (1975), the Indolence Principle by Berrendonner (1989), Lakoff and Johnson's (1980) concept of metaphorical representation, and Ranganathan's (1967) faceted system. A case study was done with a group of researchers in the field of Transport Engineering, as well as an analysis of manifestos written by a group of contemporary artists with the purpose of broadening the scope of the proposed model. We conclude that the model is feasible, although it would benefit from further studies involving different fields of knowledge.

Background
Bouramoul's (2011) paper on the semantic dimension in information retrieval, taking Query Reformulation document indexing as a starting point, led to the resumption of an investigation carried out a decade ago (Orrico, 1999) in which a theoretical model was conceived for information retrieval based on metaphors. At that time, the research was motivated by the perception that contemporaneity brought many technological conveniences for the dissemination of information. However, such conveniences have also caused a build-up of information material that demands great effort in order to appropriately handle the large number of resources on the web. This perception is becoming increasingly strong these days. Besides the quantity, another problem arises due to the ease resulting from the use of new technologies: an ever-increasing universe of users may have access to wider and more varied information, which may potentially be inappropriate for them.

Thus, this proposal seeks to resolve, in principle, two problematic situations: a) the difficulty in highlighting, among a vast set of possibilities, the information that really matters; b) limiting access to inappropriate information for some segments of the population.

Considering the assumption, as proposed by Lakoff and Johnson (1980), that the human mind represents the world through metaphors, the fields of knowledge would also be represented metaphorically by those devoted to them.

In fact, using metaphors to represent knowledge is not new. Orrico and Oliveira (2005) showed that arborescent representation gained popularity in the 16th century. Because of that, Ramon Llull’s Arbor Scientiae, written around 1300, was reedited many times, and several other representations were elaborated in different areas: grammar trees, love trees, logic trees (the Porphyrian Tree), Jesuit’s trees etc. During the Age of Enlightenment, allied to this arborescent representation, the task of organizing and classifying knowledge executed by encyclopaedia authors reveals the use of another metaphor: the world map. This is how d’Alembert describes, in the Encyclopédie, the difference between an encyclopaedia and a dictionary:

[like] a type of world map, which must show the main countries, their position and their mutual dependence, the roads that lead directly from one to another. This road is often interrupted by thousands of obstacles, which are known, in each country, only by the travelling inhabitants, and cannot be represented except in highly detailed individual maps. Such individual maps will be different entries in the Encyclopédie, and the tree, or systematic map, will be its world map. (Darnton, 1986, p. 251).
In the context of that era’s project, which relied on various strategies, knowledge thus represented was transferred to intellectuals and followed a path that resulted in the secularization of education and the emergence of academic disciplines. As a consequence of the development of specialization, the ideal of the “universal person”, whose erudition was not limited to a single field but covered all types of knowledge, was attenuated. (Burke, 2003, 82a)

Therefore, the polarization between specialization and generalization arises from the fragmentation of knowledge, a development which, at least on some occasions, was regretted. As Burke (2003, 81a) says: “General knowledge became necessary due to the connection among things and interdependence of notions, such that one piece of knowledge sheds a light on the others”.

To answer the proposed questions, we selected a group of researchers from the Transport Engineering field, whose specific object —means of transportation— is studied by various areas. Our intention was to identify how each researcher represents his/her own field of knowledge and investigation, considering that the researchers in the group come from different academic areas. Moreover, knowing that this group included researchers with varied levels of experience, we assumed that the representation of the field would be constructed during the development of activities by the group.

By means of questionnaires sent by e-mail to each group participant, we could observe the metaphoric phrases used to designate their research field. Based on the analysis of these answers and on the identification of noun phrases used to designate the field, we could see that the metaphors used by the researchers were similar. This allowed us to infer that they shared a common representational universe and to conceive the proposal for the model.

Aiming to validate the theoretical framework of the proposed model, another investigation was carried out (Coelho, 2006) in a different field, this time Contemporary Art, and particularly Conceptual Art. This study was used only to reinforce our certainty about the theoretical framework.

Objective

The objective of this text is to present a theoretical proposal for an information retrieval filter based on metaphors used in natural language.

First we present the current scenario of excessive information in circulation; then we introduce the theoretical framework that supports the arguments behind our proposal, the case study, and finally the proposal itself.

Context

As predicted by Solla Price (1965), academic production grows exponentially, doubling its production every ten years. What can be said, then, about this new space of information dissemination that is the Internet? Context shows that Orrico’s (2002) proposal, first presented for discussion in this ISKO forum during its meeting in Granada, is still valid. Raieli (2012) even points out the broader problem of aggregating documents of different natures for retrieval.

Due to Internet’s internal organization, this means of virtual communication allows a virtually unrestricted amount of data supply and access, which can cause, as pointed out by Pierre Lévy (1999), what Roy Ascott termed “second deluge”, which is characterized by the chaotic and exponentially explosive nature of telecommunications growth”.

This scenario is by no means new and information professionals seek increasingly creative alternatives, making use of natural language to enable access, search, and retrieval of information circulating on the web. Current studies of ontologies are evidence of this.
According to the initial proposal by Guarino (1998), formal ontology starts off from a specific description of the objects to be retrieved, establishing a relation of dependency between them and, according to Park (2008), the ontologies have been cited as a critical part of the projects of information systems, to a great degree because they help sustain a communication scheme in the fields of interest of a certain social group.

It is assumed that the proposal presented here can contribute to the studies of ontologies because it is part of a scenario which takes context into account in order to determine meaning. For this assertion, it bases itself on Durkheim’s (apud Burke, 2008, 78) words, according to whom ―The categories of human thought are never permanently fixed; they constantly build, unbuild and rebuild themselves; they change with place and time‖.

To achieve our broader objective, that is, helping a given group retrieve the information that interests it the most, this article is based on a theoretical semantic framework, specifically on metaphors used by certain social groups to represent both the area of knowledge in which they work and the concepts connected to it.

In order to assume that metaphors are adequate for information retrieval, this article works on the assumption that meaning is built up within an interactive-communicative process, whose organisation is governed by principles of communication. The principles underpinning our analysis are Cooperation, proposed by Grice (1975), and Indolence, by Berrendonner (1989), as shown in Orrico (1999). These principles allow us to relate Lakoff and Johnson's (1980) concept of metaphor to Ranganathan's (1967) revolutionary faceted classification system for the classification of knowledge.

**Theoretical-methodological framework**

For Paul Grice (1975), human communication is organised according to the Cooperative Principle, whose assumption is that individuals who communicate build utterances according to four maxims: Quality (Truth), Quantity (Information), Relation (Relevance) and Manner (Clarity). According to the first maxim, it is assumed that everything the interlocutor says is true; the second maintains that he/she only says what is necessary; the third, he/she only says what is relevant for communication and, finally, he/she does it in the best possible manner.

Going a step further from the previous principle, Berrendonner’s (1989) Indolence Principle establishes an interface between communication and cognition insofar as it says that in utterances there is little informative content. It is possible to formulate little informative material because the speaker takes into account that the listener already has enough information which enables him/her to infer further specifications from an utterance with underspecified content.

The interaction of metaphor with these two principles occurs because metaphor is a figure of speech that transfers a term to a scope of meaning outside its own scope in order to establish a representation of the world by means of analogies. These analogies, built within communication processes, follow the four principles presented in the previous paragraphs.

Metaphor is here understood from Lakoff and Johnson’s (1980) point of view, where they maintain that human beings organise knowledge making use of structures called idealized cognitive models [ICMs] and that categorial structures are derived from that organization. The proposal of these models admits that mental organisation occurs through the cultural construction of schemes of knowledge about the world, through the construction of what authors call ontological metaphor, which is the core metaphor from which others derive, and which help this representation. To the extent that metaphors are used to represent socially constructed categories, let us then discuss what meaning representations are.

Meaning representations refer us to the father of library sciences, S. R. Ranganathan. His idea of classification, which is both innovative and has important repercussions for the
theory of classification as a whole, reflected the modern concept of knowledge which persisted at the end of the nineteenth century and the beginning of the twentieth century, and which presupposed a precise delimitation of the area of knowledge where such concepts existed. Although nowadays we deal with imprecise limits to understanding cultural phenomena, Ranganathan’s faceted proposal builds an adequate framework within which metaphors are found.

**Case Study**

This theoretical dialogue was an important support for a case study, carried out with a team from the transport area in Brazil, consisting of twenty-eight workers with different levels of experience and from various regions of the country. Data collection was carried out by means of a questionnaire, electronically sent and answered, and its main aim for the writing of this paper was to generate representations of the transport area by analysing the following question: “What exactly do you understand by transport?”

As a methodological analysis, sentences were analysed in which the interviewee explicitly used –or left no margin of doubt about– the verbs be or represent in utterances of the type –Transport is…” or —Transport represents…”

The analysis of the replies revealed a recurrent use of terms which can place transport within three semantic fields: a) network / set / system; b) movement / dislocation; and c) manner / way / procedure.

The proposal of the theoretical model identified the relationship between this reply pattern and Ranganathan’s faceted categories, assigning to the personality category the ontological metaphor that represents the area being investigated. This metaphor was selected as ontological because it was used by most research participants and was present in the discursive context indicating the nature of the field.

Other representations appeared in the universe of answers in discursive contexts referring sometimes to the role played by the Transport field and sometimes to the conditions or manners in which it operates, with no representations related to the space and time in which Transportation occurs.

This fact indicates that, maybe because these dimensions are inherent to the Transport theme, there is nothing to be said as specific representations, and thus it is not pertinent to draw attention to them. Taking this into account, these categories were considered secondary in the creation of the filter, since, in a model intended to filter information, the focus must be on what determines specificity rather than similitude.

Based on these observations, we established a correlation among the representations of the transport field made by the researchers and the faceted categories proposed by Ranganathan. This correlation lead to Table 1, where the first column presents Ranganathan’s categories, the second presents the correlation with the proposed model, and the third presents the nuclei of the noun phrases used by the researchers to represent the field under investigation.

<table>
<thead>
<tr>
<th>Table 1: Categories of the Transport Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ranganathan’s Categories</strong></td>
</tr>
<tr>
<td>----------------------------------------</td>
</tr>
<tr>
<td>Fundamental Personality</td>
</tr>
<tr>
<td>Matter</td>
</tr>
<tr>
<td>Energy</td>
</tr>
<tr>
<td>Secondary Space Time</td>
</tr>
</tbody>
</table>
We believe that, based on the metaphorical representations, it is possible to identify those that represent the essence—or personality—of the field and the other categories, as proposed by Ranganathan, in order to create a user-oriented information retrieval filter.

This conceptual framework that supports the filter must focus on aspects of the user’s cognitive behaviour from a qualitative perspective, since it has to consider, apart from the field of knowledge in which the search will take place, the search strategies to be used and the goals of the user.

Using a filter would allow a researcher to retrieve the information s/he considers most relevant according to his/her universe of knowledge, optimising the searches and achieving a good satisfaction level.

As was mentioned, an investigation was conducted in another area, now not among scientific fields but in art. The study had the goal of extending our analysis to a different set of users in an attempt to investigate other discursive practices that could indicate inconsistencies in our model. The purpose was not to exhaust the discussion, but only to broaden the scope of the theoretical reflection.

In the art field, we turned to Contemporary Art, specifically a cultural movement called Conceptual Art. It is common knowledge that this type of art could only be appreciated by an experienced spectator, in the sense of someone who is already familiar with the symbolic universe that governs this type of art. In our study, we tried to identify the discursive construction of the metaphorical representations that the conceptual artists elaborated in their artistic essays/manifestos, aiming to reveal the representations used to establish the identity of that group.

In this group devoted to art rather than science, we again identified the regular use of metaphorical representations in the production of the group of artists, and that resulted in Table 2:

<table>
<thead>
<tr>
<th>Ranganathan’s Categories</th>
<th>Model Categories</th>
<th>Metaphors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fundamental Personality</td>
<td>Essence</td>
<td>Action</td>
</tr>
<tr>
<td>Matter</td>
<td>Function</td>
<td>Artistic expression</td>
</tr>
<tr>
<td>Energy</td>
<td>Conditions or Manner</td>
<td>War / Fight</td>
</tr>
<tr>
<td>Secondary Space</td>
<td>Space</td>
<td>Brazil</td>
</tr>
<tr>
<td>Time</td>
<td>Time</td>
<td>Military regime</td>
</tr>
</tbody>
</table>

In the specific case of this type of artistic manifestation, although Space and Time categories continue to be secondary, they need to be explicit because, in the art field, especially during the military regime, the fact of being in Brazil was something that needed to be signalled.

Based on these two studies, we believe that it is possible to broaden the reach of our model.

Conclusion

Any and all classification criteria, with their intrinsic relevance rationale, ultimately aim to meet the expectation of user satisfaction. In this case the proposal aims to use metaphors which represent categories that comply with the necessary representation conditions for this area of knowledge. This classificatory organisation aims to increase the accuracy in meeting the demands in the context of new information technologies. Metaphors in this case would help attain higher accuracy, however paradoxical this may seem at first. In fact, we regard metaphor less as a literary style feature which points towards inaccuracy and more as a degree of representation which points towards specifying concepts used in a given area of knowledge.
References