Knowledge organization of fiction and narrative documents: a challenge in the age of the multimedia revolution

Abstract
Knowledge organization (KO) research has been traditionally more oriented toward scientific documents. However, narrative, one of the oldest kinds of literature genre, is increasingly important in the multimedia culture of today, and is getting more important in applied social sciences as business administration, advertising, public relations, politics or education. This paper explores the key facets for KO in the field building on literature theory and faceted classification theory.

0: Aims and methodology
This paper presents our research on integrating the models of content analysis and knowledge organization (KO) developed in two fields: (a) library and information science – mainly using the analytico-synthetic tradition; and (b) research coming from the researchers in literature theory – mainly in the fields of narratology – the discipline that studies the structure of narrative texts – and thematology – the discipline that studies subjects as a system and their evolution.

We understand narration as a communicative strategy, deeply ingrained in the functioning of human mind and social communication, by which a sender offers – and a receiver gets – a message conveyed in a story, that is, an account of a course of events happening to a being or group of beings.
Our effort is motivated not only by our love of stories, but mainly by the growing interest in narrative information that permeates all the aspects of our contemporary culture, because of the possibilities of multimedia, a media for which narrative information is specially suitable.

In this paper, we will address this growing interest; the specificities of narratives – concretely, fictional and, among them, literary narratives; the complex community of users around this kind of information with their specific needs; and we will propose several methodological keys for an advance of KO in this important field of human information.

1: Context: the growing importance of narrative in contemporary culture
Traditionally, KO researchers – and in general information science researchers – have been more concerned about scientific documents. Certainly, this was a necessity, because information science was born precisely to deal with the explosion in scientific information. However, the revolution of today is about multimedia information, and a great deal of such information does not suit into the frame of traditional scientific reports and articles. They are much nearer to fiction and narrative documents. In fact, narrative-based communications are gaining considerable room in disciplines like education – with multimedia and video as increasing important tools in face-to-face, blended and virtual education –, psychology, and even in fields like business and management in general, where the “story-telling” movement is gaining importance in marketing, corporate culture management and political communication. The ever growing economic importance of popular culture and the correlative development of cultural industries – ever interested in the reusing of cultural artefacts and themes for new products – only adds to the importance of narratives as a tool of communication.

Our culture is reinvigorating stories as a tool of communication, and this is not strange at all, because stories are easy to remember, produce a better “closing” or Gestalt of understanding, and have a direct access to other non-logical intelligences, like emotional or aesthetical intelligence, which are being accepted – with all their virtues and faults – as a key part of human learning, or, at least by those more critical of them, as unavoidable ones.

2: Fiction and not fictional narratives
A very important part of narrative information is fiction. If a researcher considers information to be only actual facts, fiction will not be considered information. This approach can have a sound logic in science and technology, where something that it is not strictly real is a hypothesis in the best case or simply a nonsense in the rest of them. But fiction conveys cultural and – at the end – very pragmatic information, teaching people about possible courses of events, the complexity of relations… and, as it was stated before, is a medium to wrap any message to implicate the user’s non-logical dimensions, appealing to other channels of communication.

In fact, we also must remember that fiction is an artistic manifestation, present in all civilizations, and it has been studied in a scientific way since the ancient Greece, as did Aristotle in his *Ars poetica*. Thus, many researchers face this as a concrete manifestation of information, studying the structure of composition, for instance, despite the fiction tells a story on a imaginary place, as did Swift in *Gulliver’s travels*. 
Of course, there are many narrative documents that are not fiction: histories, journalistic stories, many social sciences research papers and reports... Also some fiction is not narrative – like a large part of poetry. Finally, there is a blurring frontier between fiction and nonfiction that is evident, and affects both sides of the line.

3: The specificity of fiction documents

Analysis is even more important with fictional narrative documents, because in them sense is considered implicit – in the always fragile context of the time and space of creation and reception – or even voluntarily conveyed or blurred. This is the reason because, in didactical narratives, the reader is usually briefed about the sense of the story at the end, in the form of an explanation. It is also very interesting to note that “fictionality” is a property that accepts grades, in the sense of Aristotle’s verisimilitude. Sometimes, fictional documents have been considered divided in two groups: the first conveying information and the second one, whose sole purpose would be to amuse or generate emotions (Lancaster 2003, 200). But this division should not be used to demote the importance of the second group of texts, for emotions are informational processes, very important for example in creating public opinions and ideological positions, and amusement is one of the biggest industries of our time, which, in addition, serves also the development of group and public opinion.

Moreover, stories are a way in which persons and groups explain to themselves and the others their identity and purpose. Fiction is very useful because it offers a multitude of “story forms” that we can use to learn, think and sense about possible situations that we can find upon us, near us or that we might find ourselves caught into in the future. In this sense, narration is a way to understand the world and also a way to gain knowledge about the evolution of things and persons. So, Pinker (1997) show that narrations provide a mental catalogue of the situations we can face in the future, and Hobbs (1990) see narrations as some kind of experiments where we can see or imagine how a situation might develop. In any case,

“With the multimedia revolution, a growing percentage of content is leaving the relatively unambiguous path of scientific and journalistic genres and approaching the semantic jungles of creative literature. This kind of material requires the complex kind of content analysis that is characteristic of artistic disciplines, that can be summarized in the simplified model that was proposed by the Art historian Panofsky (1955) — also applicable analogically to the analysis of fiction: description of common subjects; identification of the specific names and exemplars of these common subjects; and interpretation of the abstracts subjects that are being represented through the former.,”

(García 2006)

4: Another level of complexity: intertextuality and reception aesthetics

There is a group of fiction narratives that convey a third level of complexity. They are those texts that want to become part of a tradition, and even compete to become exemplars and reference points in it. We are talking about literature. An important part of such works is their relations with their precedent and subsequent works, and this property becomes an important part of their subjects.

Narrative documents sense is always a result of the interaction among creators, publishers and readers (reception aesthetics) throughout the times, so indexing must be open to incorporate permanently new uses and users, with their specific thematic representations at that special moment. From a practical point of view, if any narrative
text survives the occasional market, it becomes literature, and gets part of the literary tradition, and so it becomes an argument of authority, used – sometimes misused – in discussions and positions.

5: The theory of literature and subject indexing
The concept of subject in the theory of literature is quite complex. As in library and information science, the subject is the central idea around which the narration is built: children-parent relations, intercultural friendships...

A very interesting aspect for KO is the existence of thematic “universals”, which are truly cross-cultural topics that exist in different spaces and times of history. Moreover, formalists (Tomachevski 1982) detect a more analytic level of thematic analysis, with topics appearing recurrently in different moments of the narration. Topics are anyway expressed through specific and changing conceptual and aesthetical structures, called motives.

In narratology, the subject becomes very connected with the way in which it is developed throughout narrative techniques to produce a story. There are many classifications of this kind of structures. According, for example, to the relatively recent one of Tobias (1989; 1993), there are twenty master plots that are subjacent to any work of fiction: quest, adventure, pursuit, rescue, escape, revenge, riddle, rivalry, underdog, temptation, metamorphosis, transformation, maturation, love, forbidden love, sacrifice, discovery, wretched excess, ascension, and descension.

6: A model for knowledge organization of narrative documents
The most used classificatory approaches to fiction are genre, author and literature generations and movements. The last group is mainly useful for the history of literature. Author and genre usually provide a certain approximation to the emotional and intellectual experience of reading. But these approaches usually lack granularity – specificity; and, overall, they do not connect works of fiction with the information needs and desires of the readers.

In Table 1 we try to summarize the different levels of entry to the world of fiction. A general model of narrative KO should at least depart from the basic facets: personality, energy, matter, space and time. But these facets could – should? – be applied to the different levels of understanding of narrative text.

Each of these levels – and their units – has a potentiality for use among students, teachers, researchers, creative staff, authors, cultural managers, etc. Such model should therefore be contrasted with at least a prospective and extended analysis of users’ needs around fiction. The analysis of uses of fiction is a must, because information retrieval is all about satisfying users’ needs. Moreover, KO models and schemas are, in fact, produced from the target users’ needs.

The first level, literature world, is connected mainly with the needs of literature scholars, interested in the relations among authors and works, literature movements, genres, styles, narrative structures, etc. However, sub-genres can be effectively connected with the specific needs of general public, and this is probably one of the easiest ways of organizing a narrative collection by subject. It is important to remark that the big genres and subgenres do have facets. For example, a satirical treatment is possible not only in comedy, but also with novels, poetry or theatre.
<table>
<thead>
<tr>
<th>Personality</th>
<th>Energy</th>
<th>Matter</th>
<th>Space</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Literature world</td>
<td>The relation among authors and works, literature movements, genres, styles, narrative structures, etc.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Universals</td>
<td>Human relations, personality development…</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Connoted world</td>
<td>When using a story as a mirror of a subjacent one</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motives world</td>
<td>The creative expression of the denoted world</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denoted world</td>
<td>Entities, objects, persons, etc.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Potential uses</td>
<td>Different uses in education, research, creation, etc., both intended by the authors or not.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Media</td>
<td>Different communication channels</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Levels of facet analysis

The second level, universals, is very related with the needs of users. Universals codify the main approaches that users take toward narrations. This level is, indeed, quite important a topic for future research.

The level of connoted world is very important for those interested in interpretations, conveyed meaning and abstract subjects, and is a main point of interest also for the scholarship tradition in humanities. Connoted world includes, of course, intertextuality.

The world of motives is of special interest for authors of every kind and of course for cultural industries. It offers an enormous catalogue of ways of expressing a subject.

The denoted world is obviously interesting for any person searching information about objects, characters, processes, places and times. It can be treated in the same way that other informational documents are.

In sixth place, intended, potential and developing uses of a narrative document must be codified. This is especially important if we think about the interoperability of databases in the exploding Internet environment. Regarding the uses of narrative fiction, the model considers mainly six areas: educational, recreational, cultural, academic, scientific and business oriented.

Finally, multimediality must also be considered. In particular, information about creation, communication channel, reproduction conditions and conditions of use is growing in importance as the reuse of information “objects” becomes one of the main trends in the current phase of the information explosion.

7: The problem of a canonical order

However, any KO set of principles is of little interest if it is unable to produce some kind of canonical order of the world of subjects in a systematic arrangement.

Our proposal is that different schemas should be developed for each kind of users, as flexible faceted systems usually ensure. Nevertheless, we also consider very useful for general use a schema that prioritizes the hierarchy genre-subgenre-basic plots, with auxiliary facets for places, characters, ages and dates, literary movements, and narrative methodologies.
It would be very useful, too, providing users with information about the complexity of the narration, the level of fictionality and potential audiences.

8: Conclusions, recommendations and future research
Addressing narrative KO on the era of Internet means imagining models for integrating the very different uses that a narration can have. Though this information is increasingly important and widespread, many times classifications systems lack the needed granularity to ensure proper retrieval (e.g. Moraes & Guimãraes 2007). Such a model would be especially useful for the design of systems capable of facilitating the interchange of information among systems that have been designed for a specific purpose, like teaching or reusing motifs and narrative structures in film scripts. This is absolutely necessary because of the multi-meaning nature of narrative texts, which makes cooperation among many different actors necessary to achieve results.

The model we propose recognizes that there is not such a thing as a single classificatory approach to fiction, because many are the uses of fiction, and so the structure of subjects diverges from one use to the other. One person will be interested in the genre, another in the age, another in the intertextuality of a text, many in the subjacent universals that literary pieces of work illuminate and discuss, another in a character, and others in a place… On the other hand, a working canonical order is needed as a point of encounter among the different information needs that are expressed in subjects. An interesting approach in the age of the semantic Web would be to connect such alternative thematic views around a canonical one as interrelated vocabularies, in the way that the new standards, as BS 8723, make possible.

Much work must be done in clarifying each of the big levels of analysis that have been proposed in this paper. For example, as it has been lightly discussed, the world of genres, subgenres and universals is very difficult to be represented in a one-dimensional classification that can be truly useful for narrative documents discovery. It would require a complex analysis and faceting, and, so, it results to be a very interesting field for experimenting with ontologies. Also, the world of motives and figurative courses (García & García 1995; Guarido & Moraes 2009) offers very interesting insights into the not-so-obvious senses that are present in fictional discourses and poetry.

In spite of the interest of deductive methodologies, as the one here presented, there is also a need for inductive research, showing the preferences of the different types of users through well-designed user studies. One of our next steps will be a pilot study with teenagers and their teachers in secondary educations centres in Brazil and Spain.

Finally, there is a need of continuing research in how content analysis – the phase preceding actual classification and indexing – is done. The set of procedures involved in identifying, extracting and selecting contents of documents are broadly named subject analysis (Šauperl 2002; Raju & Raju 2006), conceptual analysis (Caffo 2002), aboutness (Beghtol 1986), document analysis (Gardin 1981) and content document analysis (Pinto Molina 1992). They require further research. All of them are deeply oriented by both linguistic and logical approaches (Guimarães 2009). In spite of a long tradition in cataloguing documents in libraries, it is not totally clear how the process of determining subject effectively occurs (Šauperl 2002, 1), what leads to the need of investigating its methodological dimension.
References


