NOTAS Y EXPERIENCIAS / *NOTES AND EXPERIENCES*

Construction of a taxonomy for medieval Portuguese history: problems and challenges

Filipa Medeiros*, Hermínia Vasconcelos Vilar**, José Antonio Moreiro González***

* Centro Interdisciplinar de História, Culturas e Sociedades da Universidade de Évora / Campo Arqueológico de Mértola, Portugal.
** Departamento de História, Universidade de Évora / Centro Interdisciplinar de História, Culturas e Sociedades da Universidade de Évora.
*** Departamento de Biblioteconomía y Documentación, Universidad Carlos III de Madrid.
Correo-e: fmcm@uevora.pt, hmav@uevora.pt, jamore@bib.uc3m.es

Abstract: Our main goal was to design and build a taxonomy of medieval Portuguese history using an interdisciplinary approach based on doctoral research. First, the criteria used for the selection of the vocabulary and its formal and semantic normalization were determined. Then species were listed, followed by the characterization of categories and their respective subclasses. As conclusions we highlight the successful application of the selected terms, as well as the fact that the taxonomy's categories are being continuously updated and expanded, both in their global extension and in the depth of their thematic representation. In addition we offer proposals for continuing the ontological development of this taxonomy.

Keywords: Taxonomies; KOS; information representation; medieval Portuguese history; interdisciplinarity.

Sobre la elaboración de una taxonomía de Historia medieval portuguesa: problemas y desafíos

Resumen: Se ha fijado como objetivo principal diseñar y elaborar una taxonomía sobre Historia medieval portuguesa, y hacerlo mediante un abordaje interdisciplinar y como fruto de las investigaciones de tesis. En su proceso se determinaron, primero, los criterios seguidos para la selección y normalización formal y semántica del vocabulario. Luego se listaron las especies y se caracterizaron las categorías y las respectivas subclases. Como conclusión se destaca el hecho de que se han experimentado con éxito métodos antes ya probados en otros estudios para proyectar, desarrollar y mantener taxonomías, con independencia del nivel de especificidad temática, hecho que plantea otro tipo de retos. Sin olvidarse de avanzar propuestas que le den continuidad hacia su desarrollo ontológico.

Palabras clave: Taxonomías; KOS; representación de la información; historia medieval portuguesa; interdisciplinariedad.

1. INTRODUCTION

This note describes a taxonomy aimed at organizing and representing information related to Portuguese Medieval History between the 12th and the 15th centuries, which is currently available online as a structured list. This is an applied work (Medeiros, 2014), adapting methods that have already been tested by experts in order to build the taxonomy.

The construction of a taxonomy within the scope of Medieval History is, as we shall see, a pioneering choice and, as such, there are no parameters available for comparison. It should be understood as an instrument that undergoes a continuous construction and evaluation process, but whose purposes are clear.

The construction of this taxonomy seeks to fill a gap found in the Portuguese information units specialized in Medieval History, many of which are integrated into universities and their research centres. Moreover, the construction of this controlled vocabulary seeks mainly to address the need for the indexation of digital resources, namely the ones contained in specialized databases, which are currently seen as crucial vehicles for disseminating and sharing open-access scientific knowledge. We can mention, as an example, a series of databases that have already been developed by the Institute for Medieval Studies of the Faculdade de Ciências Sociais e Humanas of the Universidade Nova de Lisboa (IEM - FCSH/UNL), available at: http://iem.fcsh.unl.pt/section.aspx?kind=noticia&id=49

2. METHODS AND RESOURCES

In order to build and develop this taxonomy we used a top-down, or general-to-specific, method (Conway and Sligar, 2002; Cumming, 2003; Gilchrist, 2003; Jagerman, 2006; Zhonghong et al, 2006; Moreiro González, 2006; Moreiro González, 2011), which is the most commonly used method for the construction of this type of controlled vocabulary. This method comprises the following stages:

a) Compiling the knowledge;

b) Reducing the number of synonyms and choosing the preferential terms;

c) Preparing a preliminary systematic structure;

d) Developing the first draft of the scheme;

e) Completing the taxonomy with all its terms;

f) Evaluating its performance;

g) Publishing its first version.

Next, we detail the specific development of each of the stages in the construction of the Portuguese Medieval History taxonomy, following the top-down method.

a) Compilation of knowledge: in this first stage we surveyed the available specialized sources of information, focusing mainly on vocabularies, lexicons and dictionaries dealing with the Middle Ages. We did not find any documents of Portuguese origin, so we collected foreign reference documents within the scope of general Medieval History that were likely to enrich the taxonomy, specifically in what regarded the definition of its general categories, both in terms of form and in terms of content. We reached a total of 11 foreign sources of information. In the absence of Portuguese sources of information, we resorted to national general histories and university textbooks of reference within the scope of Portuguese medieval studies (Serrão and Marques, 1987; Marques, 1988; Tavares, 1990; Coelho, 1991; Branco and Costa, 1992; Tavares, 1992; Serrão, 1994; Moreno, 1995; Serrão and Marques, 1996; Mattoso, 1997; Ramos et al, 2009).

b) Reduction of synonyms and choice of the preferential terms: this item corresponds to the second stage in the construction of the taxonomy, in which we set out the issues concerning the formal and semantic control of the vocabulary, according to the following analytical scheme shown in Figure 1.

It should be noted here that since the source language of the taxonomy is Portuguese, we naturally followed the guidelines provided by national standards, namely the 4036 Portuguese Standard (items 6 and 7) and the *Siporbase: sistema de indexação em português* manual (section 4, Terminology) (NP 4036, 1992; Portugal, 1998).

The morphological and syntactic control has not proved particularly complex. The Portuguese language rules were followed, as well as the national standards. The same did not occur with the semantic control of some terms, as is the case of the word "Jantar" and the term "Cantigas de escárnio e de maldizer". The first is an example of multiple meanings (semantic ambiguities via significant pathway) and the second, a case of synonymity (semantic ambiguities via meaning).

As we can see in figure 2, and according to NP 4036 (NP 4036, 1992: 13), for these cases we resorted to the use of qualifiers designated in brackets, which are intended to clarify the conceptual content to which the term refers by establishing its meaning, thus avoiding ambiguities in information retrieval.
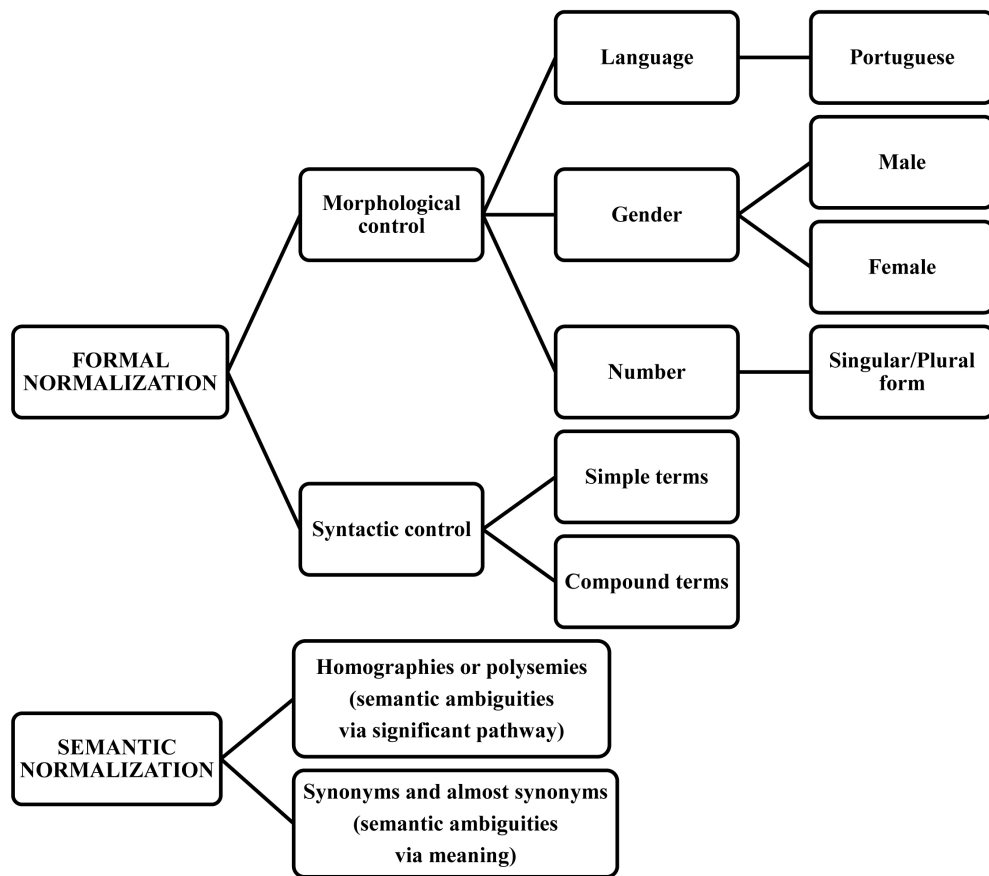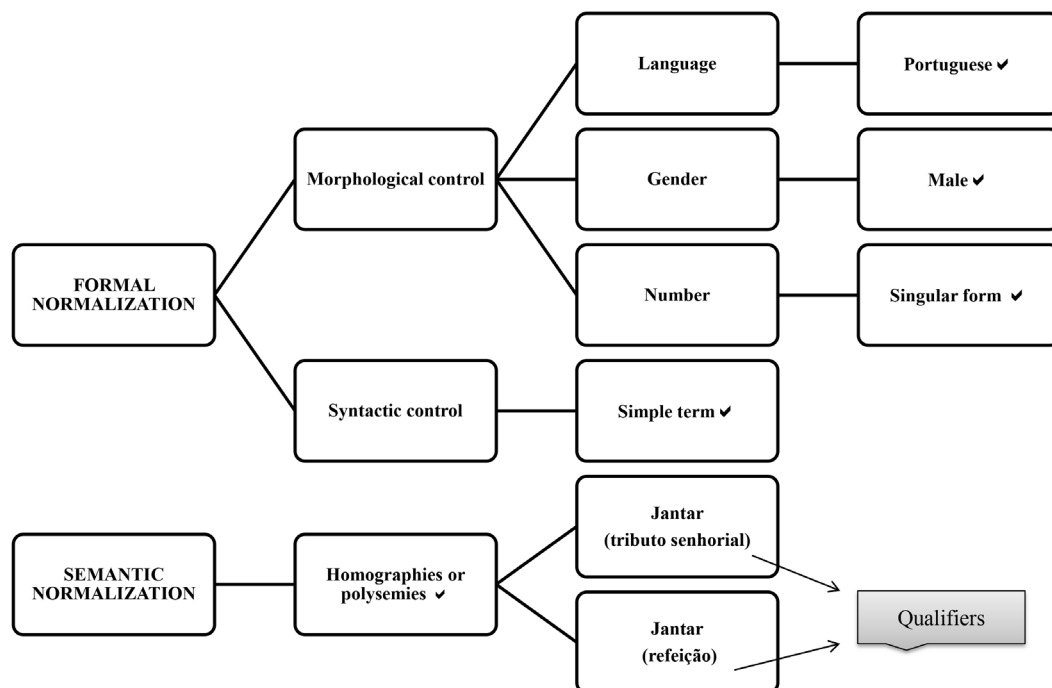
Figure 1. Stages of formal and semantic normalization

Figure 2. Formal and semantic normalization of the term "Jantar"

Figure 3 presents a case of synonymity between the terms "Cantigas" and "Canções". In this situation, our attention was focused on the term generally used by the scientific community (preferred terms), thereby facilitating the compatibility between natural language and the controlled terms present in the taxonomy.

The distribution of the terms across the hierarchical structure was made alphabetically, except in the «Events», «Personalities» and «Reigns» categories, in which the terms were introduced chronologically. This was the only way to organize them diachronically, allowing the user to browse the taxonomic structure according to the natural sequence of events, personalities and reigns, respectively:

Ex. 1128, Batalha de São Mamede

Ex. 1350-1405, Leonor Teles

Ex. 1385-1433, Reinado de D. João I

c) **Construction of the taxonomy**: After the formal and semantic normalization of the vocabulary, which followed the above-mentioned criteria and procedures, we recorded the compiled terms alphabetically, dividing them into the categories that integrated the preliminary systematic structure of the taxonomy; this was the third stage of the process.

Next, in a fourth stage, we prepared the first version of the taxonomic scheme, defining the hierarchical position of the terms within the respective categories and, consequently, ordering each specific term under the corresponding generic term. Subsequently, the terms were divided into two types of classes within each category: chains and rows. The former are vertical series of concepts and the latter are horizontal series of concepts; each series may be generic (type?) or partitive (part? whole?) (Campos e Gomes, 2007).

These procedures allowed us to move on to the fifth stage: preparing the classification superstructure of the taxonomy and filling it with all the terms. The final version includes the following 17 categories (see Fig. 4).

1) Administration

2) Fields of History

3) Historiographical sciences and techniques

4) Crown

5) Chronology

Figure 3. Formal and semantic normalization of the term "Cantigas de escárnio e de Maldizer"
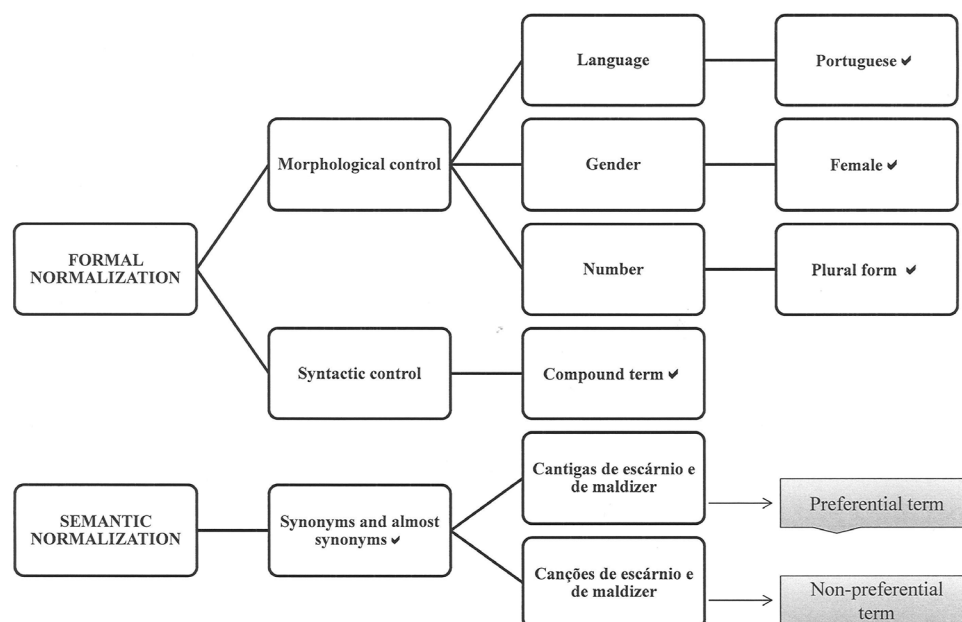

Figure 3: Formal and semantic normalization of the term "Cantigas de escárnio e de maldizer"

6) Culture

7) Economy

8) Social structure

9) Historical sources

10) Geography

11) War

12) Personalities

13) Information resources

14) Reigns

15) Religion

16) Society

17) Private life

In a sixth stage, we evaluated the taxonomy according to the recommendations of the ANSI/NISO Z39.19-2005 standard, because both the NP 4036 and the ISO 25964-2011 standards fail to mention this issue. So, we chose to conduct a heuristic evaluation of the final version of the taxonomy, since we believe that this is the most appropriate approach to the controlled vocabulary we have built; it is a domain-based taxonomy and, as such, it implies a real and permanent intervention and monitoring of the construction process by experts, which is what actually happened (ANSI/NISO Z39.19, 2005).

Therefore, we asked for the formal collaboration of two experts, one from the field of Portuguese Medieval History, and the other one from the field of Library and Information Science (LIS) –specialized in knowledge organization systems– whose evaluations we present further ahead.

In the seventh and last stage, we published the first version of the taxonomy. All the terms that had been compiled and normalized were fed into a specific taxonomy management software called Knowledge Manager (KM), marketed by the Spanish company *The Reuse Company*, which collaborates with the Departmento de Biblioteconomía y Documentación of the Universidad Carlos III de Madrid. The version we used is a downloadable test version (5.0.0). The KM allows the creation of different types of reports which can be exported in various formats (alphabetic, hierarchical or as a glossary), thus enabling terms to be fed into other applications. We decided to present the final version of the taxonomy alphabetically, insofar as, in our opinion, this option makes it possible to navigate through the semantic structure of the taxonomy, allowing an overall view of its categories and of their respective chains and rows as well.

Take, as an example, the search term "Demanda do Santo Graal". Figure 5 shows the interface of the search term in the KM and some filtering options enabled by this *software*.

In figure 6 is displayed the result of the research carried out, according to a hierarchical presentation in which we can observe the existing subordination levels between the term "Demanda do Santo Graal" and all its generic terms.

Figure 4. Overview of all categories of the taxonomy in the KM software
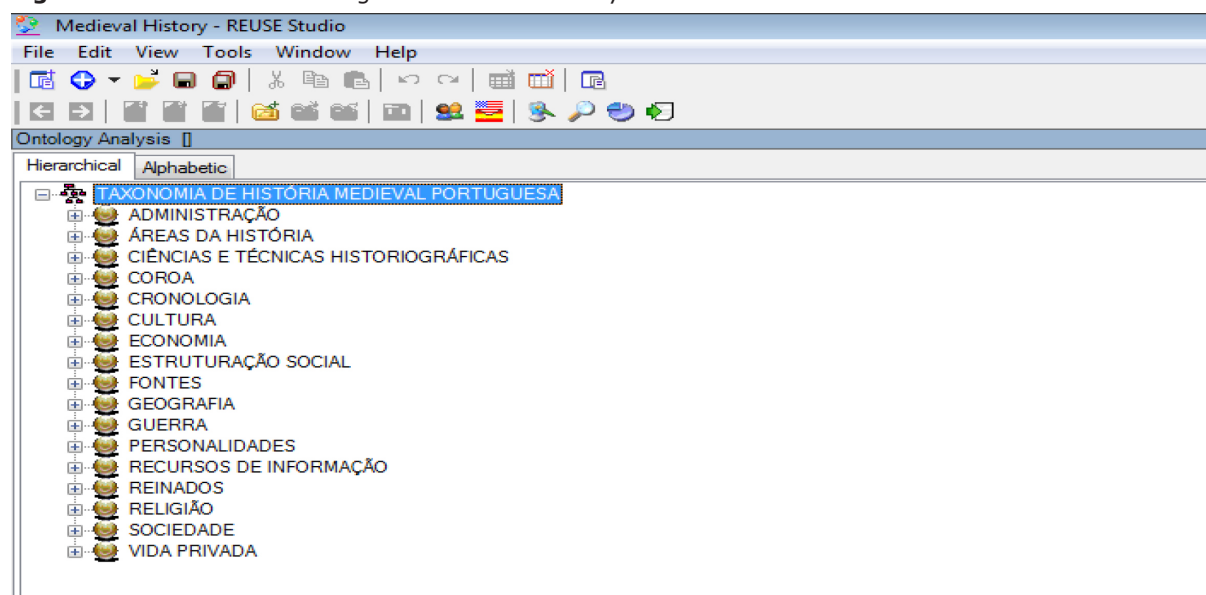
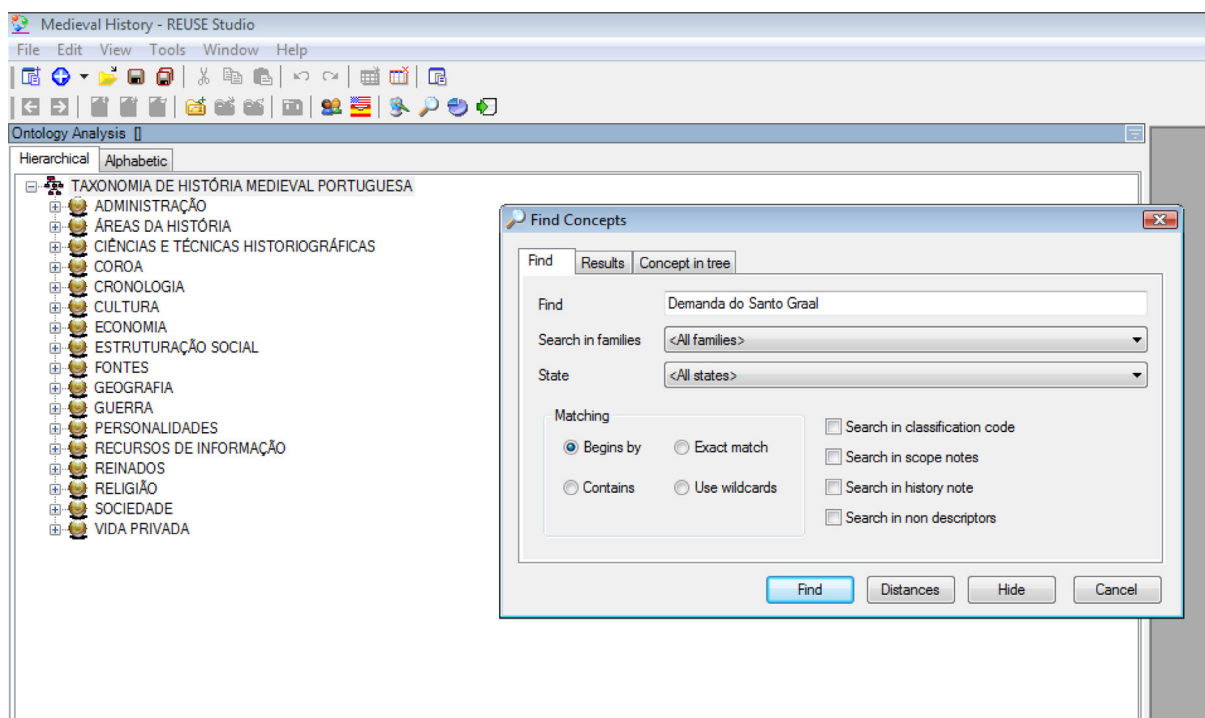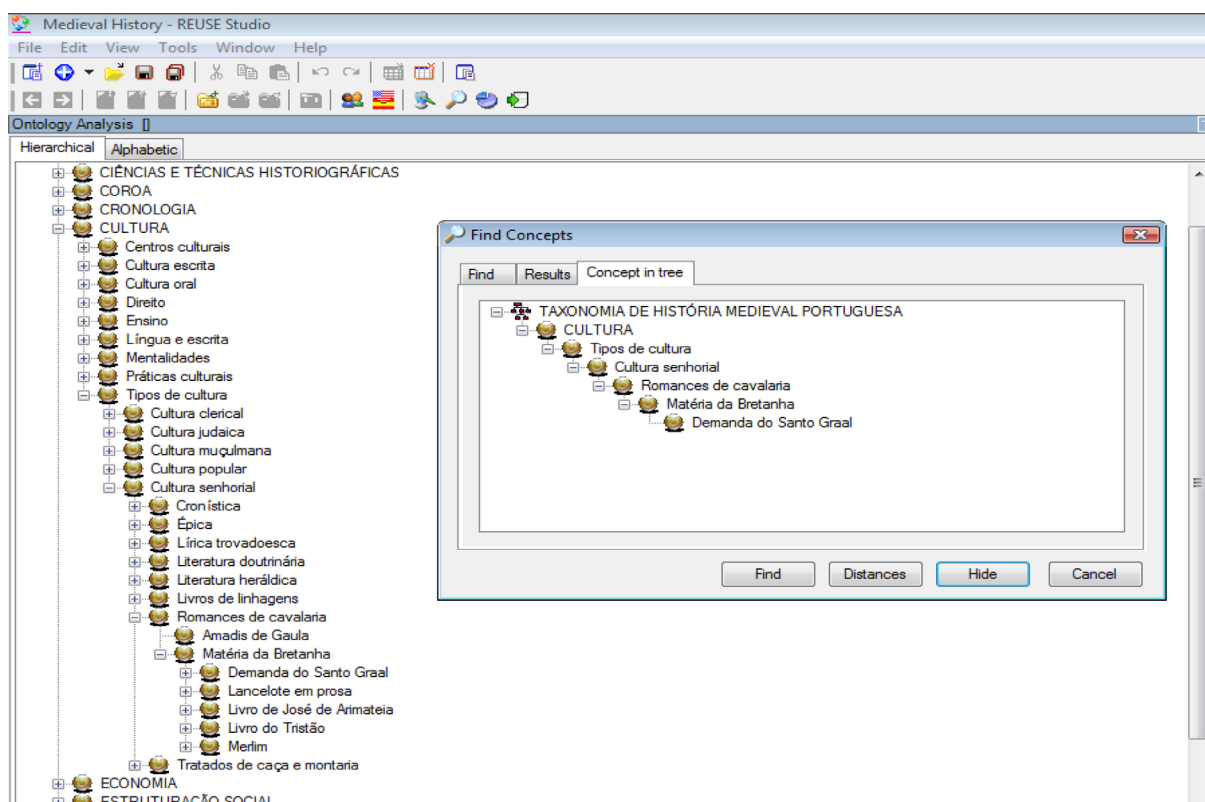**Figure 5.** Search interface in the KM software (1)

Figure 6. Search interface in the KM software (2)

3. PRESENTATION AND DISCUSSION OF THE RESULTS

Currently, the taxonomy includes 2987 terms which are constantly being reviewed and updated.

The full version of the taxonomy is available online at: http://www.en.cidehus.uevora.pt/Bases-de-Dados/Taxonomia-de-Historia-Medieval-Portuguesa

As shown in figure 7, the Portuguese Medieval History taxonomy has 17 categories, with different depths.

So, «Chronology» is the largest category, with 570 terms, while «Historiographical sciences and techniques» has just 9 terms.

The categories «Personalities», «Economy», «War», «Culture» and «Society» also stand out with 523, 486, 265, 244 and 231 terms, respectively.

On an intermediate level we list the categories «Administration», «Social structure», «Religion» and «Historical sources», which have, correspondingly, 142, 132, 114 and 93 terms.
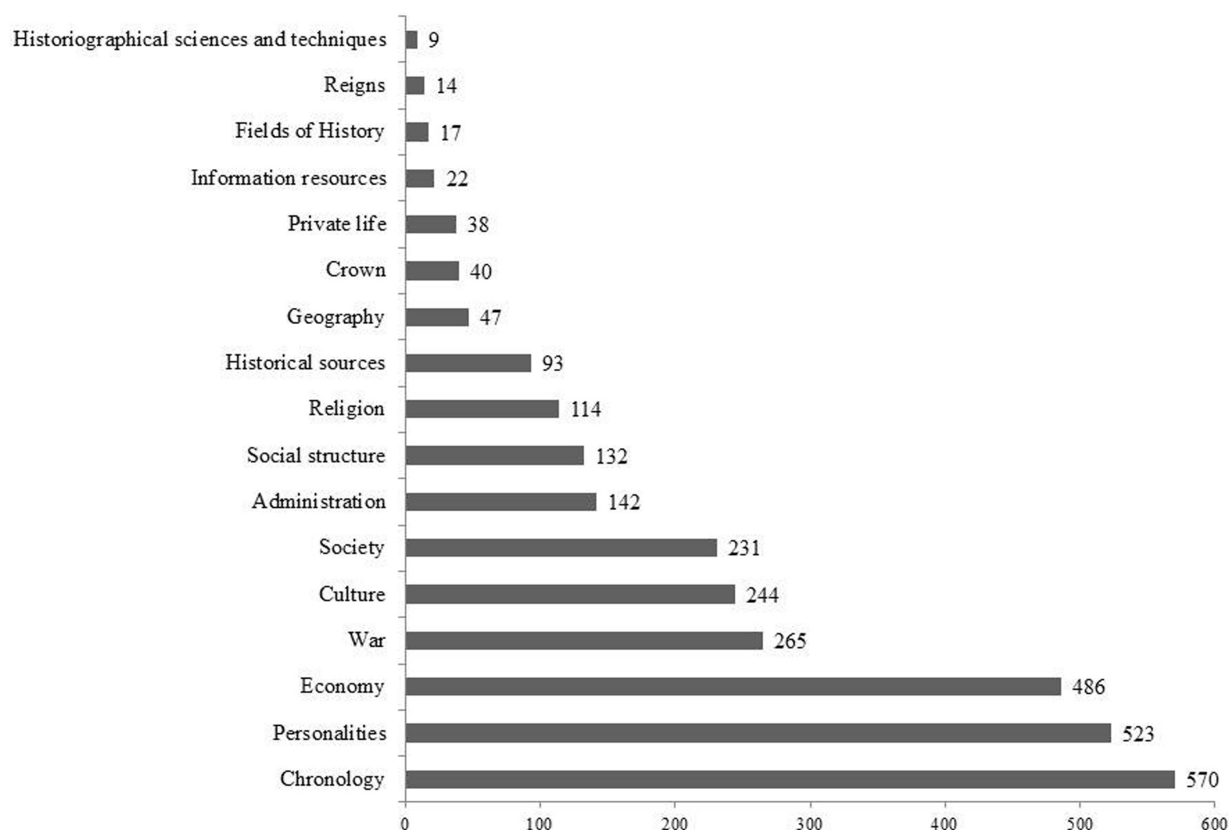
Finally, and with a smaller number of terms, we identify the categories «Geography», «Crown», «Private life», «Information resources», «Fields of History», «Reigns» and «Historiographical sciences and techniques», which have 47, 40, 38, 22, 17, 14 and 9 terms, respectively.

We have to consider the diverse sizes and depths of categories. This is due mainly to the different scopes of these specific categories. Other factors are the uneven distribution of specific sources for each category, and the variable extent to which different topics of Medieval History are represented in the scientific knowledge under evaluation - that which is developed within Portuguese universities.

With regard to the evaluation of the taxonomy, generally speaking, we asked our medievalist to conduct a semantic evaluation (regarding timeliness and communicability of the information) in addition, we asked our expert on representation of information to conduct a formal evaluation (on the morphological and syntactic aspects). For that purpose, we prepared two analysis grids, following the recommendations of the ANSI/NISO Z39.19-2005 standard. As expected in the case of interdisciplinary research works, the results of both evaluations were very positive and complemented

Figure 7. Distribution of the no. of terms across the categories of the taxonomy of Portuguese Medieval History

each other; this circumstance is essential for the work under study, given the need for a joint terminological validation.

Tables I and II show the evaluation criteria that were defined, as well as the score assigned by the experts for each criterion.

4. CONCLUSIONS

By way of conclusion, we make some essential considerations:

First, it should be noted that the taxonomy we constructed, when applied to research, successfully tested the methods previously tried by experts for designing, developing and maintaining controlled vocabularies of this nature.

We should also note that, although this taxonomy is representative of one country at one particular historical moment, it can still function as a starting point for building other controlled vocabularies with the same thematic scope, referring to different geographical locations, in the same time frame.

Finally, we stress the pioneering nature of this taxonomy within the panorama of Portuguese

Medieval Studies. Precisely due to the absence of comparative structures, both at the national and at the international levels, its construction was a great challenge that we were only able to meet thanks to an interdisciplinary collaboration between medievalists and I&D professionals. That is why we would like to stress the idea that the taxonomy we have built is, above all, a first attempt at defining a terminological categorization for this subject matter. As such, it is open for discussion and it undergoes a permanent formal and semantic evolution process, as the terms of the taxonomy are applied to existing specialized databases, on an experimental basis for the time being.

5. FUTURE DEVELOPMENTS

With regard to future developments, we consider that the taxonomy can integrate various materials from various digital resources (databases, bibliographies and other specialized compilations) into a single format (xml) and classify them all using the same taxonomy.

We would also like to emphasize that the reference taxonomy is still under construction because, even though the semantic structure is finished, we still

Table I. Evaluation grid (medievalist)

	EVALUATION CRITERIA	SCORE
1.	General and specific information sources	Very good
2.	Uniformity with regard to the representation of the various subjects included in the taxonomy	Very good
3.	Semantic relevance of the terms considering the vocabulary used by the users (timeliness of the terms)	Very good
4.	Adequacy of the hierarchical structure	Very good
5.	Efficiency of the browsing system	Very good
6.	Efficiency of the search system (retrieval and search options)	Very good
7.	Notes/Suggestions: The researchers should ask for more specific advice from experts related to the various thematic fields covered by the taxonomy.	

Table II. Evaluation grid (expert on representation of information)

	EVALUATION CRITERIA	SCORE
1.	Reference sources and procedures used in the normalization of the vocabulary	Very good
2.	Morphological and syntactic consistency of the terms	Very good
3.	Conceptual accuracy in terms of the definition of the hierarchical relationships	Very good
4.	Efficiency of the browsing system	Very good
5.	Efficiency of the search system (retrieval and search options)	Good
6.	Notes/Suggestions: The researchers should aim at diversifying and refining the available search options.	

have to develop a way to determine names through labels and schematic links which favours its reuse (Daconta et al, 2003). So, this is an essentially terminological taxonomy that does not have an ontological nature. In a second stage, we should define a naming convention aimed at enabling unambiguous identifications through the use of a formalized representation format, whose patterns should be determined according to a proper standard. This will, therefore, be the main future challenge for the taxonomy we have presented here.

6. ACKNOWLEDGEMENTS

7. REFERENCES

Bibliography on medieval history

Note: Since it is impossible to include all the bibliographic references that were used as sources of information for the selection of the terms included in the taxonomy, we decided to mention only general histories and reference university textbooks from the field of Portuguese Medieval Studies.

Branco, M. J. V.; Costa, A. M. (1992). *História da Idade Média*: caderno de apoio. Lisboa: Universidade Aberta. ISBN 972-674-370-2.

Coelho, M. H.da C. (1991). *História medieval de Portugal: guia de estudo*. Porto: Universidade Portucalense. ISBN 972-9354-07-3.

Marques, A. H. De O. (1988). *Guia do estudante de História medieval portuguesa*. 3.ª ed. Lisboa: Estampa.

Mattoso, J. (dir.) (1997). *História de Portugal: a monarquia feudal (1096-1480)*. Lisboa: Estampa, vol. 2. ISBN 972-33-1263-8.

Moreno, H. B. (1995). *História de Portugal medievo: político e institucional*. Colab. Maria da Conceição Falcão Ferreira, Luís Carlos do Amaral, Luís Miguel Duarte. Lisboa: Universidade Aberta, 2 vols. ISBN 972-674-132-7 (vol. 1). ISBN 972-674-135-1 (vol. 2).

Ramos, R.; Sousa, B.V.; Monteiro, N. G. (2009). *História de Portugal*. Lisboa: A Esfera dos Livros. ISBN 978-989-626-139-9.

Serrão, J. V. (1994). *História de Portugal: Estado, pátria, nação*. 5.ª ed. Lisboa: Verbo, vol. 1. ISBN 972-22-0271-5.

Serrão, J.; Marques, A. H. de O. (dir.) (1987). *Nova História de Portugal: Portugal na crise dos séculos XIV e XV*. Coord. A. H. de Oliveira Marques. Lisboa: Estampa, vol. 4.

Serrão, J; Marques, A. H. de O. (dir.) (1996). *Nova História de Portugal: Portugal em definição de fronteiras (1096-1325): do Condado Portucalense à crise do século XIV*. Coord. Maria Helena da Cruz Coelho, Armando Luís de Carvalho Homem. Lisboa: Presença, 1996. ISBN 972-23-2039-4, vol. 3.

Tavares, M. J. F. (1990). *Sociedades e culturas portuguesas*. Lisboa: Universidade Aberta, 2 vols. ISBN 972-674-030-4 (vol. 1). ISBN 972-674-049-5 (vol. 2).

Tavares, M. J. F. (1992). *História de Portugal medievo: economia e sociedade.* Lisboa: Universidade Aberta. ISBN 972-674-100-9.

Bibliography on representation of information

ANSI/NISO Z39.19-2005. *Guidelines for the construction, format and management of monolingual controlled vocabularies*. NISO Press: Bethesda, MD. ISBN 1-880-124-65-3.

Campos, M. L.; Gomes, H. Espanha a. Taxonomia e classificação: a categorização como princípio. *Enancib: Encontro Nacional de Pesquisa em Ciência da Informação.* [Electronic Version]. Salvador, 8, 2007. Retrieved 22 sep. 2010, from http://www.enancib.ppgci.ufba.br/artigos/GT2--101.pdf

Conway, S.; Sligar, C. (2002). *Unlocking knowledge assets*. Redmont: Microsoft Press. ISBN 0-7356-1463-6.

Cumming, M. (2003). Tomatoes are not the only fruit: a rough guide to taxonomies, thesauri, ontologies and the like. *Record Management Bulletin*, nº 113, 1-4.

Daconta, M.; Obrst, L. J.; Smith, K. T. (2003). *The semantic web. A guide to the future of XML, web services and knowledge management*. Indianapolis: Wiley.

Gilchrist, A. (2003). Thesauri, taxonomies and ontologies: an etymological note. *Journal of documentation*, vol. 59, (1) 7-18. ISSN 0022-0418. http://dx.doi.org/10.1108/00220410310457984

ISO 25964-2011. Information and Documentation. *Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval. Part 2: Interoperability with other vocabularies*. Geneva: International Organization for Standardization.

Jagerman, E. (2006). *Creating, maintaining and applying taxonomies*. Zoetermeer: E. Jagerman. ISBN 90-811-274-1-1.

Medeiros, F. (2014). *A historiografia medieval portuguesa na viragem do milénio: análise bibliométrica (2000-2010) e representação taxonómica* [Electronic Version]. Retrieved 22 oct. 2014, from https://dspace.uevora.pt/rdpc/handle/10174/11220?mode=full

Moreiro González, J. A. (2006). La representación y recuperación de los contenidos digitales: de los tesauros conceptuales a las folksonomías. In: Tramullas, J. (coord.). *Tendencias en Documentación Digital*. Gijón: TREA: 81-109.

Moreiro González, José Antonio (2011). *Linguagens documentárias e vocabulários semânticos para a web: elementos concetuais*. Salvador: EDUFBA. ISBN 978-85-232-0824-0.

NP 4036 (1992). Documentação. *Tesauros monolingues: directivas para a sua construção e desenvolvimento*. Lisboa: Instituto Português da Qualidade.

Portugal. Biblioteca Nacional. Área de Indexação e Classificação (1998). *SIPORBASE: sistema de indexação em português: manual*. 3.ª ed. rev. e aum. Lisboa: Biblioteca Nacional. ISBN 972-565-154-5.

Zhonghong, W.; Chaudhry, A. S.; Khoo, C. (2006). Potential and prospects of taxonomies for content organization. *Knowledge organization: international journal devoted to concept theory, classification, indexing and knowledge representation*, vol. 33 (3), 160-170. ISSN 0943-7444.