

# Wikipedia as a source of monolingual and multilingual information about the Spanish heritage

## *Wikipédia como fonte de informação monolíngue e multilíngue sobre o patrimônio histórico da Espanha*

María-Dolores OLVERA-LOBO<sup>1</sup>

Juncal GUTIÉRREZ-ARTACHO<sup>2</sup>

Macarena Amo VALDIVIESO<sup>3</sup>

### Abstract

The aim of the study was to analyze the online encyclopaedia Wikipedia as a tool for the dissemination of monolingual and multilingual information about the Spanish heritage. The sample consists of catalogued buildings and sites of cultural interest in a Spanish region. For that purpose, we examined to what extent articles about the buildings and sites are represented in Wikipedia that is, what kind of information is dedicated to each of them. Three different types of search in the encyclopaedia (browsing, lists and internal search engine) were performed for several months. In addition, we established to what extent Wikipedia includes multilingual information regarding the sample, that is, the number of languages in which the information is expressed, as well as articles translated into languages other than Spanish. The results show that only 21.73% of the buildings and sites that constitute the final sample have a separate article in Wikipedia, of which only 20.85% are in several languages, and 50.00% are properly described as translations. This leads us to conclude that, given the collaborative nature of this tool and in view of the cultural and economic importance of historical and cultural heritage, institutions should be responsible for promoting the multilingual dissemination of this type of information.

**Keywords:** Cultural goods. Historical heritage. Multilingualism. Wikipedia.

### Resumo

*Este trabalho analisa a enciclopédia online Wikipédia como uma ferramenta de difusão de informação monolíngue e multilíngue sobre patrimônio histórico da Espanha. A amostra é composta por bens imóveis de interesse cultural catalogados de várias províncias espanholas. Para o efeito, estudou-se em que medida estes bens estão representados na Wikipédia através de um artigo próprio, ou seja, com informação independente dedicada a cada um deles. Procedeu-se à realização de três diferentes tipos de pesquisa na enciclopédia (navegação, por meio de anexos e através do motor de busca interno) durante vários meses. Por outro lado, determinou-se o grau em que a Wikipédia inclui informação multilíngue em relação à amostra, identificando o número e tipos de línguas nas quais se representa essa informação, bem como os artigos presentes noutras línguas diferentes do espanhol que são o resultado de uma tradução*

<sup>1</sup> University of Granada, Department of Information and Communication. Colegio Maximo de Cartuja, s/n., 18071, Granada, Spain. *Correspondência para/* Correspondence to: M.D. OLIVEIRA-LOBO. E-mail: <molvera@ugr.es>.

<sup>2</sup> University of Granada, Department of Translation and Interpreting. Granada, Spain.

<sup>3</sup> University of Granada, Faculty of Documentation and Communication. Granada, Spain.

Received on 3/3/2015, resubmitted on 9/8/2015 and accepted for publication on 9/2/2016.

*propriamente dita. Os resultados mostram que apenas 21,73% dos bens que constituem a amostra final possuem um artigo independente na Wikipédia, destes, somente 20,85% se encontram em várias línguas, sendo 50,00% traduções propriamente ditas. Isto leva-nos a concluir que, dada a natureza colaborativa desta ferramenta e dada também a importância cultural e econômica do patrimônio histórico, as próprias instituições deveriam responsabilizar-se em promover a difusão multilingue deste tipo de informação.*

**Palavras-chave:** Bens culturais. Patrimônio histórico. Multilinguismo. Wikipédia.

## Introduction

The Web is a well-established means for disseminating information. It is not subject to geographical boundaries and has an even more numerous audience. Due to that nature, above all easy access from any part of the world, there is a need to offer multilingual information which is capable of meeting the needs of a wide range of users, thus overcoming language barriers. This is the context in which Wikipedia has come to be one of the main sources of information for Web users. According to its own definition, it is 'an encyclopaedia, understood to be a means for the compilation, storage and transmission of structured information. A wiki can be edited by anyone and it is free (WIKIPEDIA, 2015). In effect, it is a free encyclopaedia, that is, there are no legal restrictions on use or distribution, and it can be modified. It is online (on the Web), collaborative (created by the online community), and multilingual (it aspires to include all the languages possible). It contains information in almost 300 different languages, of which English is the most representative with more than 4 million articles. A large part of the work on Wikipedia is done collaboratively, seeking to reach a degree of consensus. There is space for debate that facilitates coordinated participation, such as the Community Portal, the Village Pump, and Wikiprojects.

The present study analyzes the online encyclopaedia Wikipedia as a tool for the dissemination of monolingual and multilingual information about the Spanish heritage. The sample consists of catalogued buildings and sites of cultural interest in a Spanish region. For that purpose, we examined to what extent articles about the buildings and sites are represented in Wikipedia that is, what kind of information is dedicated to each of them. Three different types of search in the encyclopaedia (browsing, lists and internal search engine) were performed for several months. Furthermore, we reviewed the organisation of information in categories and subcategories in the encyclopaedia. In addition, we

established to what extent Wikipedia includes multilingual information regarding the sample, that is, the number of languages in which the information is expressed, as well as articles translated into languages other than Spanish.

The results show that only 21.73% of the buildings and sites that constitute the final sample have a separate article in Wikipedia, of which only 20.85% are in several languages, and 50.00% are properly described as translations. This leads us to conclude that, given the collaborative nature of this tool and in view of the cultural and economic importance of historical and cultural heritage, Spanish institutions should be responsible for promoting the multilingual dissemination of this type of information.

### Wikipedia, source of multilingual information

The information on Wikipedia is organized, basically, into categories and articles. A category can include subcategories and links to articles. It can, in turn, belong to a parent category. In order to correctly categorize the entries or articles, a series of principles are followed, for example, such as not being redundant and including the article in all the categories to which it may belong due to its nature. However, problems can arise when carrying out this organization due to the application of the Unicode standard – rules for encoding characters designed to facilitate the electronic processing of the information.

To create a standard article, Wikipedia offers a guide for the standardization of the contents in the encyclopaedia. The guide recommends the inclusion of an introductory section or concise definition of the concept or subject of the article. The main body of the article must be structured into sections and subsections and be accompanied by sections for notes and references, bibliographical references, additional bibliography, and a section for external links. Finally, the

categories in which the article is indexed must be stated, along with the interlanguage or interwiki links, that is to say, the links that connect articles on the same subject in different languages. They may be translations of the article or newly written texts. Translations in Wikipedia are not generated by an integrated automatic translation tool but rather are done by contributors who, we assume, have sufficient knowledge of the languages concerned. Nevertheless, although users are encouraged to provide correct translations in the language(s) in which they have language skills, this does not prevent the occasional instance when a contributor uses an automatic translator and simply 'cuts and pastes', thereby incorporating low quality translations into the encyclopaedia. The pages or profiles of numerous contributors show signs, symbols and awards, among other things, which give information about language skills and other details of interest.

The rapid growth of Wikipedia in recent years has led to numerous studies that use the resource in different settings. Responding to the academic interest promoted by the access to multilingual information and natural language processing techniques involved in these processes (PETERS *et al.*, 2004), Wikipedia has shown itself to be a resource for different multilingual applications (SORGRA; CIMIANO, 2012) such as translation of queries (NGUYEN *et al.*, 2009), extraction from bilingual dictionaries (YE *et al.*, 2012), automatic translations and the creation of comparable corpuses (KIRAN *et al.*, 2012; DREXLER *et al.*, 2014), multilingual search for answers (OLVERA-LOBO; GUTIÉRREZ-ARTACHO, 2011), among many other contexts.

In effect, the most noteworthy aspect of Wikipedia is that it is considered a source of multilingual information. Indeed, what is known as Wikipedia consists in truth not of one sole wiki, but rather a large number of wikis, each distinct and independent according to the language, thus forming a multilingual system. It follows that the contents, management and users of each Wikipedia are distinct and independent.

There are currently more than 280 Wikipedias according to the language in which the information is edited. The analysis presented herein focuses on the Wikipedia in Spanish, which according to its number of content pages is ranked tenth among the different Wikipedias. As of 1st October 2014, it had 1,128,727 valid

pages of content (a measurement that had been taken in September 2013 revealed the existence of 1,050,076 pages of content, showing an increase of 78,651 articles, about 7% in 13 months).

The general objective of the present research was to analyze the representation of the information on Wikipedia about buildings and sites of cultural interest in Spain, and the multilingual processing of that information. It must be borne in mind that heritage and the way it is managed undoubtedly influences activities such as tourism and economic development (FERNÁNDEZ TABALES; SANTOS, 1999; FERNANDO; LÓPEZ GUZMÁN, 2004). While it is expected that there are topics that receive great multilingual treatment in the different Wikipedias due to the global interest they generate, we consider that analyzing the multilingual character of Wikipedia through local information makes it possible to determine whether this information source fulfils its aim of being a multilingual tool for Web users. When we refer to local information, we do so in the sense that it is linked to a specific geographical place, but is also of general interest as it forms part of the heritage, which is capable of generating attention and instilling the desire to become familiar with it and/or to visit it in a great number of Wikipedia and Web users.

Other specific goals arise from the general objective are addressed in the section on the research goals, namely: (a) to know to what extent Andalusian buildings and sites of cultural interest are represented in Wikipedia by way of their own articles, that is, with separate information dedicated to each; (b) to analyze how the information is organised in the encyclopaedia relating to the subject of study; (c) to determine to what extent Wikipedia includes multilingual information regarding the sample, identifying the number of languages in which that information is represented; and (d) to confirm which of the articles in languages other than Spanish are the result of translations.

### Methodological procedures

The study sample consists of buildings and sites, and activities of ethnological and cultural interest in the provinces of Granada and Huelva contained in the *Catálogo General del Patrimonio Histórico Andaluz*

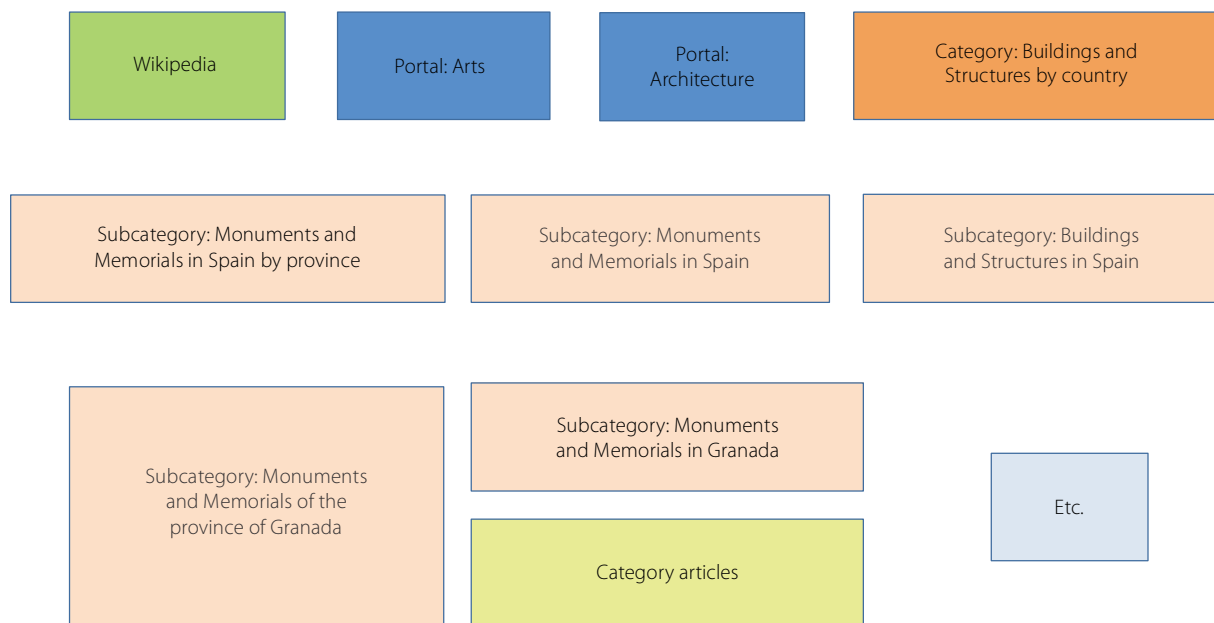
(CGPHA, General Catalogue of Andalusian Historical and Cultural Heritage), which is the responsibility of the Department of Education, Culture and Sports of the Regional Government of Andalusia. The catalogue is a tool used to safeguard the registered buildings and sites and facilitate inquiries and promotion. The two provinces in the Autonomous Community of Andalusia, Granada and Huelva, were chosen because they have the largest and smallest number, respectively, of buildings and sites of this type of heritage indexed in the catalogue. With the aim of establishing how the information about the sample is represented in Wikipedia, different types of search have been performed. Furthermore, the subject categories used by the encyclopaedia to organise the information have been analyzed. Similarly, we determined how many of the buildings and sites in the sample provide information in various languages, and which of them are translations. The procedure is described below.

In order to know how many of these buildings and sites are described in a separate article in Wikipedia, dedicated exclusively to the building or site in question, three types of search were performed for one year (2013-2014):

a) Browsing through the categories established in Wikipedia related to the sample, for example, “Monuments and memorials of the province of...” and “Defensive towers of...”, among many others. In order to perform the search, we first visited the categories, starting from the most generic (Portal: Arts) to the most specific at the final level. Figure 1 shows an example of navigation through the categories in Wikipedia.

b) Analysis of the lists related to the buildings and sites of cultural interest that compose the sample. These lists, included in the categories previously analyzed by the browsing method, proved to be of interest given that, on the one hand, they helped to shed light on some buildings and sites that had gone unnoticed in the prior procedure and, on the other hand, they revealed a number of mistakes on Wikipedia that were worth recording.

c) Queries through the search engine integrated into Wikipedia made it possible to locate, using the name of the building or site in question, those that had not been found through the two previous routes/procedures. This search engine only retrieves information included in Wikipedia in Spanish since, as stated above, each Wikipedia is independent. This method allowed the



**Figure 1.** Example of navigation by browsing (2014). Source: Author (2014).

retrieval of some elements that were inadequately categorized and were not correctly included in the corresponding list.

While searching for the various buildings and sites in the encyclopaedia, a number of problematic cases were detected. This was due to the manner in which they had been registered in Wikipedia or even in the Catalogue from where they had been obtained. Similarly, there were mistakes made by users when creating the encyclopaedia's content. Faced with such cases, the following criteria were adopted:

a) Historical buildings included in more than one municipality: some buildings and sites appear listed more than once, given that they are included or pertain to more than one municipality. For the purpose of the analysis they were taken into account only once.

b) Buildings and sites with uncertain representation in the encyclopaedia: in some cases, the buildings and sites whose information is analyzed do not have their own separate entry in Wikipedia; however, a certain amount of information about them is included in the article corresponding to the municipality to which it belongs. In order to record these cases, the category called 'uncertain representation' was created. However, these have not been included in the analyses together with those buildings and sites that are represented in Wikipedia in a separate article.

c) Coincidences in some of the names of the buildings and sites: there are buildings and sites included in the CGPHA that coincide with at least one of the alternative names that are assigned and listed in the section "other name". In that event, given that it does not really concern different buildings and sites, if in Wikipedia there is a separate article for that "other name", that is the entry taken into account for the purposes of our analysis.

d) Errors of omission of buildings and sites: two types of omission have been detected: 1) catalogued buildings and sites of historical and cultural heritage that do not appear in Wikipedia, and 2) buildings and sites present in Wikipedia that do not appear in the CGPHA. In neither of those cases the buildings and sites were part of the analysis.

e) Consolidation of buildings and sites: in certain cases, what appears in the catalogue as one building or site, as in the case of "The Alhambra and the Generalife", appears in Wikipedia as two separate articles, one "The Alhambra", the other "The Generalife". In such cases, they have had to be considered as two distinct buildings or sites.

f) Duplicated: sometimes there is more than one article in Wikipedia about the same building or site. This occurs because some buildings or sites have different alternative names in addition to the official name, causing confusion. In that event, only one of the names was taken into account, choosing the article that provides more information.

After searching the Wikipedia for buildings and sites of historical and cultural interest that constituted the initial sample, and having applied the previously described information selection criteria, the final sample consisted of 971 buildings and sites, among which a wide variety of elements was found, ranging from buildings of international popularity to subaquatic spaces unknown to the public at large.

The search for buildings and sites of cultural interest in accordance with the various procedures described above revealed that of the 971: 1) only 211 (21.73%) of the buildings and sites of the sample were found to be represented in a separate Wikipedia article during the data compilation period; 2) 719 (74.05%) were not represented in Wikipedia during the data compilation period; 3) 41 (4.22%), being the remainder, were deemed to be of uncertain representation. They cannot be included in the represented group as they did not have a separate article, but there was considerable information about them in the article in which they had been included.

Therefore, the analysis focused on the 211 buildings and sites represented in Wikipedia with a separate article or entry, containing the following information:

– Method of retrieval: it was recorded whether the information relating to the particular building or site was found through browsing, via the corresponding list or by using the internal search engine of Wikipedia.

– Position in the structure of Wikipedia: the categories and/or subcategories that nest the buildings and sites are shown.

– Name of the building or site: in some cases more than one name has been included; when the name that appears in Wikipedia is an alternative one, or similar, rather than the official name listed in CGPHA.

– Municipality where the buildings or sites are located: this makes it possible to locate them geographically and avoid the possibility of confusing them for other buildings and sites in other places with the same or similar name.

– Date of creation: shows the date the article was created and, in case it is written in more than one language, it is possible to know which one was created first.

– Last revision date: makes it possible to identify if the information is up to date.

– Languages in which it is available: interlanguage links in the article.

– Translations: when available in more than one language, it indicates which are translations of the Spanish article.

## Results

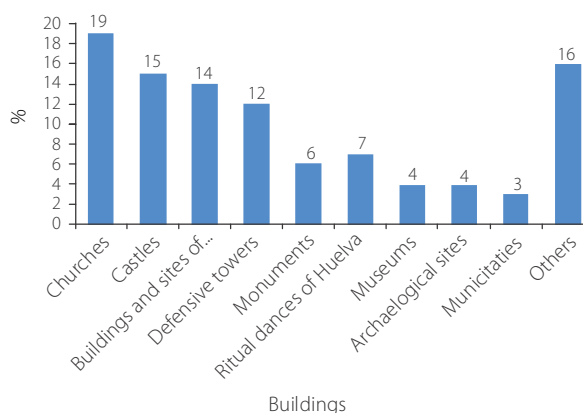
### Method retrieval

Of the 211 articles that composed the sample analyzed, the vast majority (184 articles, 87%) had been found in Wikipedia through browsing given that it was the first method of information search applied with the aim of checking the structure of the topic categories and subcategories. Thus, as expected, a smaller number of articles was retrieved when searching through the list (15 articles, 7%) – the second method used – and, finally, through the search engine (12 entries, 6%), since no new search for the buildings and sites previously found through browsing was performed. However, when applying the latter two retrieval methods, several articles were found that would have been otherwise missed. Indeed, if the categorization of the articles was perfect and there were no human error, all the articles would have been retrieved through browsing.

### Categories in Wikipedia

In the *Catálogo General del Patrimonio Histórico Andaluz*, from which the sample was obtained, the buildings and sites analyzed were found to be organised around categories such as: activities of ethnological interest, sets of historical buildings, historical gardens, places of ethnological interest, places of industrial interest, monuments and memorials, historical sites, archaeological areas, heritage areas and other non-classified areas. Wikipedia provides a greater quantity and variety of categories and subcategories for structuring contents of different topics. Historical and cultural heritage is no exception. As there is a greater diversity of categories in the encyclopaedia, each of them has a fewer amount of elements. Indeed, there are several Wikipedia categories where only one article associated with our sample was found to be nested, as seen in Table 1. The category “Buildings and sites of cultural interest” in the Table includes all types of buildings and sites from our sample, which have been categorized in Wikipedia at that generic level and it does not have a more specific category linked to them. Figure 2 shows the categories with more than five elements in the encyclopaedia. The other items have been included in a single category called ‘others’.

The results show that, of the buildings and sites included in the study sample, churches and castles are the most widely represented in Wikipedia. In agreement



**Figure 2.** Main buildings and sites of the sample according to their category in Wikipedia (2014).

Source: Author (2014).

with the information in the CGPHA, both of those categories would be within the much more generic category 'monuments and memorials', which includes a larger number, more than 60%, of the buildings and sites in the sample.

### Multilingual articles and translations in Wikipedia

The interlingual links of the 211 buildings and sites in the sample have been examined to check if there was an article about the building or site in a language other than Spanish. The amount of articles available in more than one language was very small (44 entries; 20.85%), with the large majority of articles only available in Spanish (167; 79.15%), although some specific entries have a large variety of interlingual links available, as is the case with the Alhambra, whose fame and worldwide recognition has undoubtedly contributed to that fact. Nevertheless, it cannot be ruled out that in some cases it is possible that there were articles in other languages included in different Wikipedias which did not include their corresponding interlingual link.

As seen in Table 2, of the 44 articles that are available in more than one language, 22 (50.0%) are in two different languages, nine in more than five languages, five in three languages, five in four languages and three in five different languages.

The language other than Spanish in the sample with a larger number of multilingual articles is Portuguese (85.8% of the total of the entries), 36.4% in English, 34.1% in French and 27.3% in Catalan. Figure 3 shows the most representative languages of the multilingual entries analyzed.

As shown, not all the interlingual links refer to translations, some refer to a newly prepared entry. We found which the multilingual entries are translations. Accordingly, we detected that the language into which the original Spanish articles have been most frequently translated is Portuguese (16), followed by English (5), French (3), Catalan (1) and Galician (1). One reason for the greater number of translations into Portuguese other than languages is due to the high productivity of some of the contributors who translate Wikipedia articles. A review of the multilingual entries of our sample (44 in

**Table 1.** Buildings and sites of the sample according to their category in Wikipedia (2014).

Wikipedia category	n° of buildings and sites	%
Churches	40	18.96
Castles	32	15.17
Buildings and sites of cultural interest	29	13.74
Defensive towers	26	12.32
Monuments	14	6.64
Ritual dances of Huelva	14	6.64
Museums	8	3.79
Archaeological sites	8	3.79
Municipalities	6	2.84
Baths	5	2.37
Hermitages	4	1.90
Convents	3	1.42
Bridges	3	1.42
Chapels	2	0.95
Palaces	2	0.95
Monumental gates	2	0.95
Dolmens	2	0.95
Archives	1	0.47
Libraries	1	0.47
Cathedrals	1	0.47
Bastions	1	0.47
Parks and Gardens	1	0.47
Bullrings	1	0.47
Historic sites	1	0.47
Royal chancery	1	0.47
Botanical gardens	1	0.47
Paleontological sites	1	0.47
Andalusian coast	1	0.47

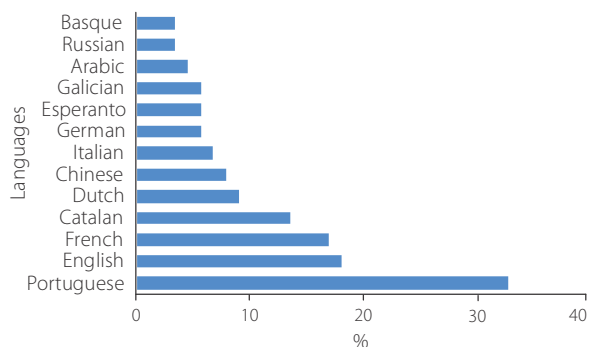
Source: Author (2014).

**Table 2.** Buildings and sites of the sample in more than one language in Wikipedia (2014).

Languages	N	%	% of the total
2	22	50.00	10.43
3	5	11.36	2.37
4	5	11.36	2.37
5	3	6.83	1.42
More than 5	9	20.45	4.26
Total	44	100.00	20.85

Source: Author (2014).

total) shows that entries that are translations (22), the most productive author's first language is Portuguese and, according to the author, he or she has also a good knowledge of English, French and Spanish. This user translated 14 of the 26 translations found.



**Figure 3.** Most representative languages of the multilingual information of the sample (2014).

Source: Author (2014).

## Conclusion

Given the inevitable limitations of the study sample and the dynamic nature of information in Wikipedia, the results of the analysis are temporary/limited. However, we consider that this does not prevent us from drawing a number of useful conclusions. Upon completion of the study, it can be confirmed that regarding our subject of interest, namely the subject domain of historical and cultural heritage and the characteristics of the sample as previously outlined, Wikipedia's scope is somewhat inadequate. The percentage of representation reaches only 21.73% of the buildings and sites of the sample and furthermore, many of the entries and articles included in the encyclopaedia contain only sparse information about the element dealt with, and not only in other languages but also in Spanish. However, in many cases it is clear that the building or site concerned did not lend itself to the compilation of further information about it. A feature that undoubtedly influences representation is the fame or international recognition of the buildings and sites, those with greater popularity are present or better represented in Wikipedia.

In addition, it has been found that the encyclopaedia's complex framework of categories is occasionally counterproductive when it comes to retrieving information. The majority of mistakes found in the encyclopaedia are related to the linking of content, but these errors are not as numerous as one might have thought as this seems to be the consequence of rapid growth in content and answering speed and updating.

It can also be confirmed that using interlingual links within an article it is possible to access the corresponding entries in other languages pertaining to other Wikipedias, given that these links act as bridges assisting interaction between the different Wikipedias. Here again, only entries for buildings and sites acknowledged to be of international interest have a considerable amount of interlingual links, such as Alhambra, the *Capilla Real* (Royal Chapel) of Granada, and some historical centers in certain municipalities but, in general, the multilingual information about these aspects of Spain's historical and cultural heritage is limited?

Finally, the examination of the Wikipedia articles with interlingual articles indicates that a number of them are translations performed by users, although not all the articles in different languages that are about the same subject are necessarily translations. Some contributors prefer to create a new article in another language.

In sum, going beyond the polemic caused by the quality of the information contained in Wikipedia (SAORÍN, 2012), it is beyond the question that it is a remarkably useful and popular tool. However, much more could be obtained from it, above all from the perspective of retrieving multilingual information, the perspective under consideration in our study. The sparse scope in the encyclopaedia regarding this type of information, of crucial importance from a cultural and economic point of view, allows us to conclude that, given that Wikipedia is a collaborative resource, it would be desirable for institutions to show greater interest in promoting the dissemination of information about historical and cultural heritage.

## Acknowledgements

This research was conducted within the framework of the project HAR2012-38562 "Platform of analysis and access to multilingual web information about historical and cultural buildings and sites in Spain", funded by the Spanish Ministry of Economy and Competitiveness, Government of Spain.

## Contributors

All authors participated in all phases of the research article.



## References

- DREXLER, J.; *et al.* Wikipedia-based corpus for contextualized machine translation. In: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION. 9., 2014, Reykjavik. *Proceedings...* Reykjavik: ELPA, 2014. p. 3593-3596.
- FERNÁNDEZ TABALES, A.; SANTOS, E. Turismo y patrimonio histórico en Andalucía: magnitudes generales y estrategias de planificación. *Boletín de la Asociación de Geógrafos Españoles*, v. 28, p. 119-134, 1999.
- FERNANDO, L.; LÓPEZ GUZMÁN, T. J. El turismo como motor de desarrollo económico en ciudades patrimonio de la humanidad. *Pasos: Revista de Turismo y Patrimonio Cultural*, v. 2, n. 2, p. 243-256, 2004.
- KIRAN, K. N.; SANTOSH, G. S. K.; VASUDEVA, V. Multilingual document clustering using Wikipedia as external knowledge. In: INTERNATIONAL CONFERENCE ON MULTIDISCIPLINARY INFORMATION RETRIEVAL FACILITY, 2., 2012, Vienna. *Proceedings...* Vienna: Springs, 2012. p. 108-117.
- NGUYEN, D.; OVERWIJK, A.; HAUFF, C. WikiTranslate: Query translation for cross-lingual information retrieval using only Wikipedia. In: CROSS-LANGUAGE EVALUATION FORUM CONFERENCE ON EVALUATING SYSTEMS FOR MULTILINGUAL AND MULTIMODAL INFORMATION ACCESS, 9., Corfu. *Proceedings...* Berlin: Springer Berlin, 2009. p. 58-65.
- OLVERA-LOBO, M. D., GUTIÉRREZ-ARTACHO, J. Language resources used in multi-lingual question answering systems. *Online Information Review*, v. 35, n. 4, p. 543-557, 2011.
- PETERS, C.; BRASCHLER, M.; CLOUGH, P. *Multilingual information retrieval: From research to practice*. New York: Springer, 2004.
- SAORÍN, T. *Wikipedia de la A a la W*. Barcelona: UOC, 2012.
- SORGRA, P.; CIMIANO, P. Exploiting Wikipedia for cross-lingual and multilingual information retrieval. *Data & Knowledge Engineering*, v. 74, p. 26-45, 2012.
- WIKIPEDIA. *A translation of the Spanish language version*. Available from: <[http://es.wikipedia.org/wiki/Wikipedia:La\\_encyclopedia\\_libre](http://es.wikipedia.org/wiki/Wikipedia:La_encyclopedia_libre)>. Cited: Feb. 8, 2015.
- YE, Z.; HUANG, J. X.; HE, B.; LIN, H. Mining a multilingual association dictionary from Wikipedia for cross-language information retrieval. *Journal of the American Society for Information Science and Technology*, v. 63, p. 2474-2487, 2012.

