Project Acronym: Europeana v2
Grant Agreement number: 270902
Project Title: Europeana Version 2

D7.7: Midterm Report on Innovative Multilingual Information Access

Revision	Final
Date of submission	31.12.2012
Author(s)	Juliane Stiller, Humboldt-Universität zu Berlin Maria Gäde, Humboldt-Universität zu Berlin Vivien Petras, Humboldt-Universität zu Berlin
Dissemination Level	[Public]

Project co-funded by the European Commission within the ICT Policy Support Programme

REVISION HISTORY AND STATEMENT OF ORIGINALITY

Revision History

Revision No.	Date	Author	Organisation	Description
Draft 1	07.12.2012	Juliane Stiller, Maria Gäde	Humboldt-Universität zu Berlin	First draft version
			WP7-Management group	First comments on strcuture
Draft 2	14.12.2012	Juliane Stiller, Maria Gäde	Humboldt-Universität zu Berlin	Second draft
	20.12.2012	Antoine Issac	WP7-Management group	Final comments
	21.12.2012	Vivien Petras	Humboldt-Universität zu Berlin	Chapter 7 and revision
Final version	21.12.2012	Juliane Stiller	Humboldt-Universität zu Berlin	Final

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise.

Contents

Executive Summary

The deliverable reports on the research conducted during the first 15 months in task 7.4 Multilingual Access to Content. It reports work on developing novel or alternative interaction models for multilingual access to Europeana.

The focus of this report is to determine multilingual access features in digital libraries and especially in cultural heritage digital libraries. An analysis of existing information systems in the GLAM-domain was conducted to establish and collect existing solutions for searching and browsing multilingual content. In a next step, Europeana was studied with a focus on implemented multilingual interactions. Challenges in implementation and recommendations on how to overcome them are given. Based on this, possible interaction models focused on user-assisted translation features are discussed and solutions presented evaluating the pros and cons of such an approach.

The survey of cultural heritage websites showed that many sites deal with multilingual issues such as users from different countries and objects in different languages. Nevertheless, multilingual access is mainly limited to offering the interface in several languages. Only in rare cases this is extended to the metadata of the objects. For Europeana, this means that in terms of multilingual access it can be a trailblazer guiding the direction for future developments. It was shown that Europeana already offers many multilingual access points. Major achievements are the multilingual enrichments of the metadata that facilitate retrieval across languages and the curated exhibitions, which highlight content in several languages. To improve these features and offer seamless multilingual access, some challenges need to be faced. Some are easy wins that can overcome confusion on the user side by providing more help texts. Others intervene with the search workflow introducing more clicks and cognitive efforts on the user side.

The report is organized as follows; chapter 2 provides the theoretical background information about the different levels of multilinguality in digital libraries. Chapter 3 presents an introduction into user interaction models and user-assisted translations. Chapter 4 provides an overview of current functionalities provided by cultural heritage websites derived by a content analysis of 32 websites. Solutions on the display of multilingual data and searching, browsing and exploring multilingual content are given. This analysis guides the investigation of existing multilingual access components implemented in Europeana and their limitations, which are presented in chapter 5. Chapter 6 exemplifies a use case of an alternative multilingual interaction workflow - user-assisted query translation and result presentation. The report concludes with the evaluation of ranking improvements for Europeana and suggests future work.

At the beginning of the project, an adaption of the deliverables and milestone was arranged. There will be two main deliverables reporting on the three subtasks (7.4.1 Novel user interaction models for multilingual access to Europeana, 7.4.2 User-assisted translation, 7.4.3 Leveraging user-driven & multilingual semantic data for enhancing Europeana object metadata), this mid-term report on innovative Multilingual Access (M15) and a final report on innovative multilingual access (M29). In alignment with this change, all tasks last until the end of the project (29 months).

1. Introduction

This deliverable reports work on developing novel or alternative interaction models for multilingual access to Europeana. Within task 7.4 Multilingual Access / Translation, use cases will be developed which guide the design and usability of novel interaction models supporting effective multilingual access to cultural heritage objects. Of particular interest in this task are collaborative features, which can be leveraged to improve translations and to enrich metadata with content in new languages.

Figure 1 illustrates the three main pillars guiding the development of interaction models for accessing multilingual content. Enriching provider content multilingually focuses on multilingual metadata enrichment during the ingestion process. These enrichments can be user-driven or automatically added. Collaborative features to enhance translation consist of features, which allow users to add translations to given terms in the metadata. The last strand is multilingual user-driven data, which comprises the leveraging of user-assisted query translation or social tagging. The interweaved strands inform the development of use cases and interaction models focusing on how to present this multilingual content to users.

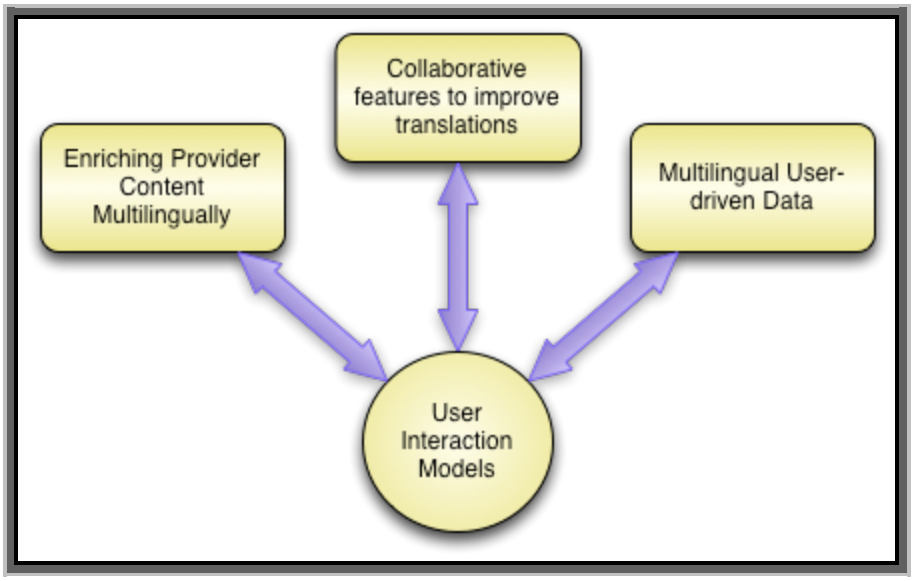

Figure 1: Sub-tasks in 7.4 and their interconnection

In this deliverable, the basis for developing these interaction models for accessing multilingual content and enriching content multilingually is laid out. The focus of this report is to determine multilingual access features in digital libraries and especially in cultural heritage digital libraries. An analysis of existing information systems in the GLAM-domain was conducted to establish and collect existing solutions for searching and browsing multilingual content.

In a next step, Europeana in particular was studied with focus on implemented multilingual interactions. Challenges in implementation and recommendations on how to overcome them are given. Based on this, possible interaction models focused on user-assisted translation features are discussed and solutions presented evaluating the pros and cons of such an approach. Goal of this deliverable is to lay the basis for developing concrete use cases and mock-ups on how collaborative features and multilingual data presentations should look like and can be used. This will be reported in the final report on innovative multilingual access to content (month 29).

The report is organized as follows; chapter 2 provides the theoretical background information about the different levels of multilinguality in digital libraries. Chapter 3 presents an introduction into user interaction models and user-assisted translations. Chapter 4 provides an overview of current functionalities provided by cultural heritage websites derived by a analysis of 32 websites. Solutions on the display of multilingual data and searching, browsing and exploring multilingual content are given. This analysis guides the investigation of existing multilingual access components implemented in Europeana and their limitations, which are presented in chapter 5. Chapter 6 exemplifies a use case of an alternative multilingual interaction workflow - user-assisted query translation and result presentation. The report concludes with the evaluation of ranking improvements for Europeana and suggests future work.

2. Dimensions of Multilinguality in Digital Libraries

Language diversity is the barrier to accessing and exploring content in the web and digital libraries (Large and Moukdad, 2000). To overcome this, digital libraries need to offer multilingual access to content on multiple dimensions. From a system point of view, multilinguality is either implemented through an interface language adaption or through access to content in different languages. The user side mainly deals with the issues to retrieve and explore content, which is not in the user's preferred language or which he cannot understand at all.

Three main levels of multilingual access in information systems – multilingual interfaces, multilingual search and browsing and multilingual result representation and translation - can be described and possible implementations presented. A detailed description and discussion of multilingual access features provided by Europeana can be found in chapter 5.

2.1 Multilingual Access in Digital Libraries

Multilingual Interface

The localization or internationalization of interfaces is the basic level of multilingual information access, sometimes referred to as MLIA. The customization of the interface according to the user's preferred or native language assures that users can access and understand a website irrespectively of their origin. Usually, system designers and stakeholders agree on either user-assisted or automatic interface language change options. Automatic interface language changes are realized through cookies that store information about the users' origin or preference based on the language of the browser or the country the user is coming from using his geo-location information. User-assisted language changes can be either provided via drop-down menus, buttons or flags. Some systems use a combination of both approaches. Europeana for example offers a drop-down menu presenting all available languages to the users. At the same time a language cookie is set once a user showed a language preference and is stored and remembered for future visits.

Multilingual Search and Browsing

Multilingual information retrieval to content is usually referred to as cross-language information retrieval (CLIR). CLIR allows users to find documents in several language irrespectively of the query language (Oard and Diekema, 1998) To overcome the language barrier between the query language and the object language different solutions have been applied so far (Oard and Diekema, 1998, Oard, 1998).

The most common approach is the query translation. The query translation process requires several steps, which are either performed automatically or with assistance by the user. A first step is the identification of the query language. Some systems require the user to determine the query language and target language whereas other systems perform hidden multilingual search. Once the source language is identified the query needs to be processed and translated into the target language(s). If the collection contains documents in different languages the query needs to be translated into all available target languages. Due to the short length of queries language identification is not a trivial task. Especially ambiguous terms can lead to wrong translations. Systems that support user-assisted query translation usually present translation candidates from which one can choose the most appropriate ones. Depending on the number of languages it could be useful to ask the user to determine the target language(s) beforehand.

Another approach for multilingual search is the translation of documents into all languages supported by a system. It allows transferring the translation from retrieval time to indexing time. Research showed downsides to document translation such as time and storage consumption and maintaining translations over time (Peters et al., 2012). This especially applies to systems supporting many different languages. In this case, the "interlingua" approach is another solution. It uses one language as pivot between source and target language. Chapter 6 provides a detailed outline of the use case user-assisted query translation.

Multilingual browsing allows users to access content and discover objects they were not aware of. Especially if users do not speak the languages the objects or their metadata are in, it is essential to offer browsing capabilities that support serendipity and discovery of the unknown. The implementation depends on the provided browsing options, like classifications of facets. The translation of concepts in different cultures adds another complexity to the translation process.

Multilingual Result Representation and Translation

The representation or even translation of retrieved objects from different countries in several languages poses another challenge for multilingual information systems. Systems can either display all results in one merged list or separated by language. Users should be able to refine search results according to their preferred language(s). Another opportunity would be that the users indicate into which language the results should be translated in advance.

For textual objects, it needs to be determined if a full translation is desired or if metadata translation is sufficient. If objects are available in several languages, the display of the multilingual data is a concern. The main questions here are:

- Which criteria is used to decide the display language?
- How can the user switch between different languages?

In search result representation, it is essential to offer support for users to enable them to determine the relevance of objects that are not in their language. Understanding translation alignments helps users to reformulate their queries and assess the relevance of the retrieved objects. The MultiSemCor web interface experimented with different presentations of bi-lingual corpora providing the users with translation alignments on word and sentence level (Ranieri et al., 2004). This web interface is mainly targeted on browsing text corpora but the presentation of the different alignment levels can also be transferred to search results.

To effectively support multilingual users in retrieving information in languages they might not understand, information systems need to offer means to help users through the information seeking process. Table 1 aggregates functionalities for the different support tools systems can offer for multilingual users (Peters et al., 2012 p. 96).

Support tools	**Functionalities for multilingual users**
Query formulation support	Query translation (e.g., language selection, select/deselect translated terms, back translation of query terms)
Evaluation support - document selection and examination	Provide summary of results (e.g., present results written in different languages, generate and translate document surrogates) Translate selected document
Query reformulation support	Edit query translation (e.g., query expansion and translation refinement)
Browsing support – collection and results	Multilingual controlled vocabularies and classification schemes

Table 1. Cross-language functionalities (Peters et al., 2012, p 96, Table 4.3)

2.2 Multilingual Users

Another dimension of multilinguality in digital libraries is represented by the cultural and linguistic diversity of users. Users that want to access multilingual content might have different language backgrounds and skills. Monolingual users need higher query translation support as well as more full translation opportunities in order to evaluate results. Users that can read and/or write at least one foreign language might want to find results in different languages without the need for full translation services.

The needs and expectations of the heterogeneous user groups with different cultural and linguistic backgrounds accessing cultural heritage information systems should be taken into account when designing multilingual systems. This goes beyond the localization of the user interface and might entail different browsing and searching habits of users coming from different linguistic backgrounds.

3. Interaction Models and User-driven Translations

3.1 Interaction Models and Multilingual Interaction Models

Interaction is an interdisciplinary term that is used across domains with slightly different meanings. Here, the definition from Human-Computer-Interaction is used which includes all interaction between a human and a computer. In this field, different aspects are applied to further determine scope and purpose of interactions. Considering the goal of an interaction or the domain the system is tied to, this research field analyses the inherit problems which occur when users interact with the information system. “HCI involves the design, implementation and evaluation of interactive system in the context of the user’s tasks and work” (Dix, 2004)[p.4]. Closely related to it are *interaction design* and *information architecture*. The first one deals with designing means for interaction with a digital product (Cooper et al., 2007). Solving issues related to navigating to the right information and providing the user with a clear structure on how to access the information defines *Information architecture.*

Another pillar in the domain of Human-Computer-Interaction is the analysis of tasks a user has to successfully fulfill to reach a certain goal. This is commonly referred to as *task analysis* which is an integral part of designing the user experience on a website. Interaction patterns are the user-friendly and successful solutions which support the user to fulfill these tasks and are developed to solve common web problems such as logging into an account (Crumlish and Malone, 2009) [p.10].

The influence of culture on Human-Computer Interaction is becoming a more and more relevant research area especially as websites want to serve users with diverse language and cultural backgrounds. A meta study of journal papers in the field found different streams, researchers were following: Culture-HCI frameworks, display design, effect of culture and localized interfaces (Clemmensen and Roese, 2010). The differences in cultural and information behavior and retrieval were not the focus of these studies.

In the information seeking literature, the influence of using a user’s native or other languages in search tasks are also studied. For example, Ford & Gelderblom examined cultural differences in interface design using Hofstede’s cultural dimensions (Hofstede and Hofstede, 2001) in order to identify their impact on the performance in human-computer-interaction based on the measures accuracy, speed and performance level of users (Ford and Gelderblom, 2003). Interface designs, which match the cultural dimension pole of the user, were compared with the opposite pole of each dimension. The results showed no sufficient evidence that these cultural dimensions had significant influence on human performance, although the accommodation of high uncertainty avoidance, masculinity, collectivism and high power distance increased the usability of interfaces for all users.

Ghorab et al. investigated the influence of different linguistic or cultural backgrounds using multilingual search logs and found that the information seeking behavior differs significantly (Ghorab et al., 2010).

Kralisch et al. examined the impact of users’ cultural backgrounds on information behavior. For the study of web interfaces, results showed significant cultural influences on navigation patterns in terms of time, linearity of information access, and amount of accessed information (Kralisch et al., 2005).

A similar study was conducted with Arabic children using the International Children’s Digital Library (ICDL). The study revealed differences in the information seeking behavior of Arabic

children to children from other language backgrounds in the same system. It also acknowledges the fact that users with specific language backgrounds have specific and varying user needs (Bilal and Bachir, 2007).

For task 7.4, interaction models and patterns for multilingual access to Europeana will be developed. This inherits workflows to design tasks such as the multilingual enhancement of object metadata or user-assisted translations. It also entails the display of multilingual data and how to simplify the user interface for users speaking different languages and coming from different countries.

3.2 User-Assisted Translation

User-assisted Query Translation

The multilingual search process requires iterative interactions between the system and a user. Figure 2 illustrates the cross-lingual search process highlighting the cognitive effort that is connected to repeated decisions within a multilingual task.

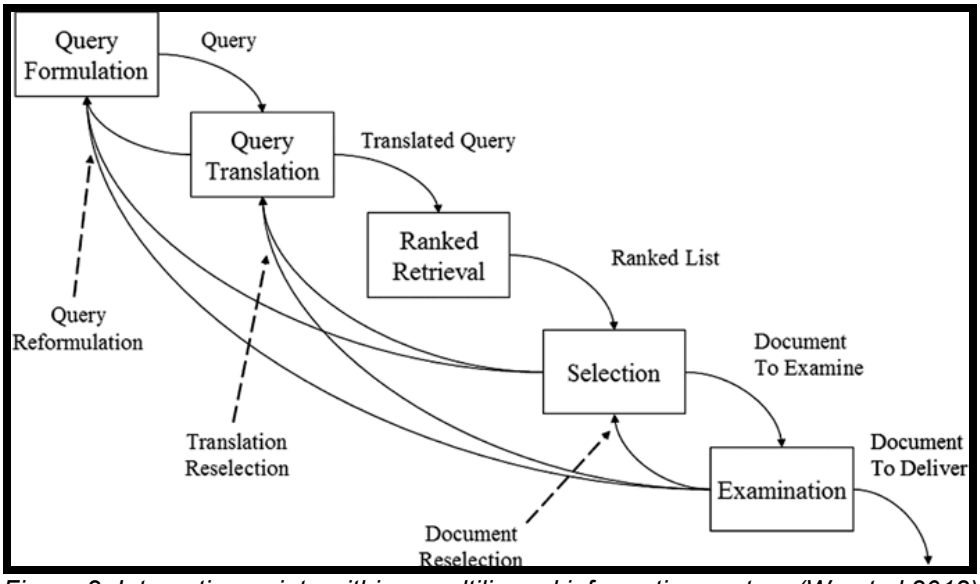

Figure 2. Interaction points within a multilingual information system (Wu et al 2012)

User-assisted query translation approaches try to improve the translation process by leveraging user input. Different point of views concerning the implementation and use of user-assisted query translation exist in the literature. While some studies highlight the importance and effectiveness of user-input for the retrieval process, others point out the challenges of user-assisted translation within systems that support many languages (Petrelli et al., 2003).

Research within this field deals with three main questions (Oard et al., 2008)

1. How should user-assisted query translation be implemented?
2. How are users interacting with user-assisted query translation features?
3. Does user-assisted query translation improve search results?

In line with interactive cross-language information retrieval research, several approaches for leveraging user input have been presented, including query suggestion, expansion, disambiguation and relevance feedback (Wu et al., 2012), (Gao et al., 2007). User input can be harvested from log files analyzing search queries, user-generated tags, annotations and result click history.

Using search logs from the European Library[1] Bosca and Dini examined the occurrence of translation pairs within search terms (Bosca and Dini, 2009).

Some interactive CLIR systems were developed to study the use and effectiveness for different use cases. An experiment with the interactive MLIA system ICE-TEA (Interactive Cross-language search Engine with Translation Enhancement showed that relevance feedback features improve the retrieval outcomes (Wu et al., 2012). The study could not find any difference between the single effectiveness of query expansion and translation enhancement or the combination of both.

Through a number of studies with The Maryland Interactive Retrieval Advanced Cross-Language Engine (MIRACLE) the impact of user-assisted translation on search strategies was examined (Oard et al. 2008).

Within the Cross Language Evaluation Forum (CLEF), several interactive retrieval experiments have been conducted. In the context of iCLEF, automatic translation settings have been compared to interactive translation with regard to retrieval improvement. He et al. 2003 (He et al., 2003) found that users were more successful in finding relevant documents when they could use contextual information such as synonyms and example sentences. Although highly interactive translation processes can improve search results it has been stated by a variety of studies that the high user effort is only required when the automatic translation fails (Petrelli et al., 2003). Furthermore it was observed that users feel more comfortable selecting phrasal suggestions than inspecting foreign query terms (Lopez-Ostenero et al., 2005). In general context information via external sources like Wikipedia is helpful to select appropriate translations (Gonzalo et al., 2008).

User-assisted Translations - Crowdsourcing

The translation could be also user-generated with an active call for participation. This is also called crowdsourcing and entails outsourcing translations traditionally made by professional translators to a large undefined audience (Howe, 2010). Research in this area is scarce and a further analysis of the topic with regard to implementation on cultural heritage sites can be found in chapter 4.5.

3.3 Multilingual & Semantic Object Metadata Enrichment

Next to translating the query to enable cross-lingual search, it is important to provide the object metadata in several languages. Not only does this support retrieval but also browsing tasks in unknown languages and therefore the exploration of unknown items. There are several ways to enrich objects multilingually:

a. Automatically enriching metadata
b. Leveraging user input through log files, social tagging, etc.
c. Crowdsourcing multilingual metadata though games etc.

Automatically Enriching Metadata

Automatically enriching metadata adds translations or controlled vocabulary to the objects' metadata, which allows the object to be found in languages that differ from language the metadata in.

Europeana and related projects have a long tradition in researching methods for multilingual semantic enrichment. It is also referred to as multilingual or semantic tagging (Isaac, 2010), (Isaac, 2011). In work package 2 of the EuropeanaConnect project, a whole task was

[1] http://www.theeuropeanlibrary.org/tel4/

[2] http://www.galateas.eu/

dedicated to the multilingual mapping of vocabularies to enable cross-lingual browsing and search. Results of this search are now implemented in the production site of Europeana multilingually enriching metadata fields (Boer et al., 2011). The quality of these enrichments plays a crucial role on the impact of these enrichments. A preliminary study with Europeana data revealed flaws and offers a framework for ensuring qualitatively high enrichments which improve the user experience (Olensky et al., 2012).

Furthermore, content can be enriched by automatically expanding metadata with external resources such as Wikipedia links. Here, one needs to be aware that cultural bias can be introduced to the digital objects based on cultural differences inherited in the use of certain terms (Callahan and Herring, 2011).

Leveraging User Input

Another approach is to leverage the user input and harness user interactions with the system to find translation candidates and improve existing dictionaries or letting translation work done by users. Two different types of user-driven data can be distinguished here:

a) User, data which is created by users interacting with the system, e.g. queries or result clicks;
b) User-generated content such as social tags, annotations and comments.

There are several methods to aggregate this data:

a) Log files;
b) Mapping user-generated tags or annotations in different languages.

In an information system with users from different linguistic backgrounds the potential amount of queries in different languages is quite high. Finding translation pairs in query logs of information systems is one goal of the EU-funded project GALATEAS[2] building on the algorithms developed in CACAO[3]. Based on the assumption that users type the same query in several languages in a multilingual information system, they developed an automatic approach to aggregate these translation pairs (Bosca and Dini, 2009). The results are promising and GALATEAS will offer a similar service for digital library administrators.

Additionally, there is the possibility to map potential language equivalents in social tags to enrich existing dictionaries or metadata. A study on social tags' potential to bridge language gaps concluded that power tags in different languages happen to be translations of each other. Nevertheless, cultural differences have an influence on the choice of tag (Eleta and Golbeck, 2012).

[2] http://www.galateas.eu/

[3] http://www.cacaoproject.eu

4. A Content Analysis of Multilingual Interactions in Digital Cultural Heritage

An overview of multilingual access features and novel interaction models currently used within the cultural heritage domain was derived from the analysis of 31 cultural heritage websites. The list was aggregated from scanning mailing list, domain-specific journals and websites dealing with digital cultural heritage. As many institutions only offer monolingual content and metadata only a limited number of websites could be examined with regard to multilingual features.

4.1 Sample Websites

Table 2 shows the types of sites occurring in this sample: definitions were developed on the basis of a survey initiated by OCLC on social metadata (Smith-Yoshimura and Shein, 2011). The definition of museum, archive, library and community sites is influenced by this referenced survey analysis; the other categories were developed based on the requirements of this purposeful sampling.

Site type	Definition	Example	# of sites
Museum	Websites providing access to the resources of a museum and visitor information.	http://www.louvre.fr	11
Archive	Websites providing access to the resources of an archive and visitor information	http://www.nationaalarchief.nl/	3
Library	Websites providing access to the resources of a library and visitor information	http://www.perseus.tufts.edu/hopper	3
Aggregator	Websites offering a single access point to the resources of several institutions or organization. Here the affiliation to a certain region or type of organization is main characteristic.	http://www.europeana.eu	7
Collection	Websites offering a single access point to resources that are united by a theme retrievable in the content.	http://www.philaplace.org	4
Community	These websites are living from and for the content of the user and the community. They can be arranged around a theme or a specific topic.	http://historypin.com	3

Table 2. Sample websites

The location of websites across different countries of origin is shown in figure 3. This is mostly based on the location of the hosting institution.

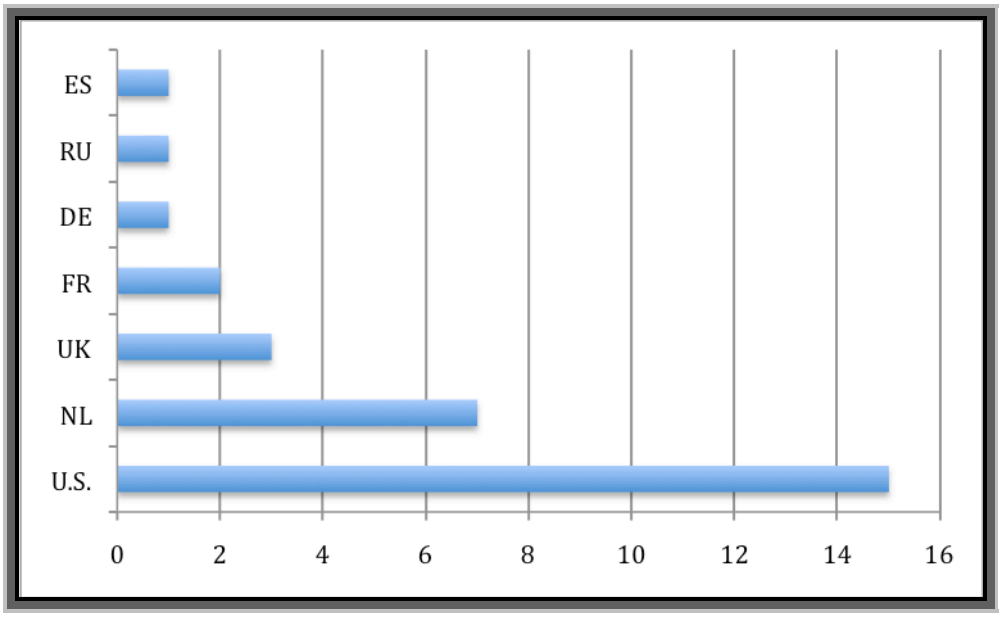

Figure 3. Location of websites

Roughly half of the analysed websites originate in Europe, the rest is located in the U.S. The majority deals with content in different languages as cultural heritage collections are often linguistically diverse. The describing metadata is rather monolingual depending on the hosting institution and the anticipated audience for the content. Nevertheless, to gain greater reach many websites also offer their metadata in one more language, mainly English. Figure 4 shows how different institutions handle multilingual offerings of their metadata. Most of the websites provide their metadata only in one language. This is mainly the case for smaller projects and organization originating in the United States. If the institution is known across borders and has international significance such as the Google Art project, metadata is likely to be in several languages. In some cases the metadata language depends on the language the digital object was provided in. This is for example the case for Europeana. On some websites the metadata language also changes with the interface language chosen by the user. Only one website, the International Children's Digital Library, offers truly multilingual metadata with the goal to make its content accessible across countries.

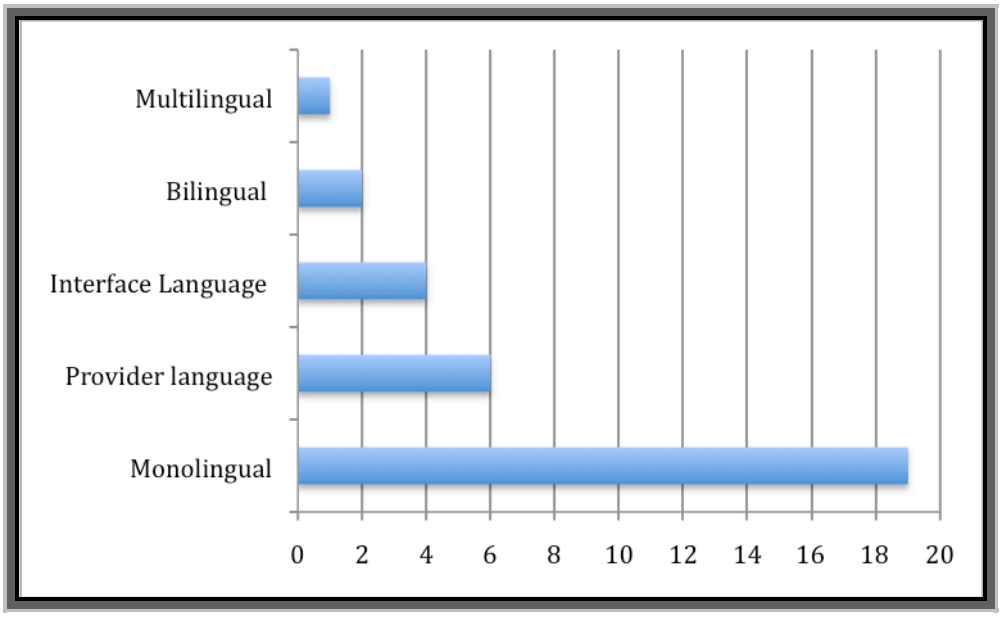

Figure 4. Language of metadata content

The most common observed multilingual feature implemented is the interface language change. Almost all sites from Europe support such a feature but only a couple of sites coming from the U.S. Some of the websites used third party solutions for the interface translation such as Google Translate[4] other relied on in-house translation.

The following sections discuss different multilingual features offered by cultural heritage websites in more detail looking at flaws and successes.

4.2 Multilingual Display

A prerequisite for multilingual access to content is offering not only multilingual objects such as books in different languages but also the describing metadata in different languages. Currently only a few information systems exist providing multilingual metadata descriptions. Multilingual representation is mainly implemented at the multilingual interface level. Through localization, the content of all static pages can be translated into the users' preferred language. In most cases, this does not include the translation of metadata. The translation of metadata is a cost- and labour-intensive procedure and rarely happens before the actual ingestion. Figure 5 shows a screenshot of the Archives Portal Europe[5] which shows the navigational features translated into Lithuanian whereas the metadata in the case findings aids are in the language they were provided in.

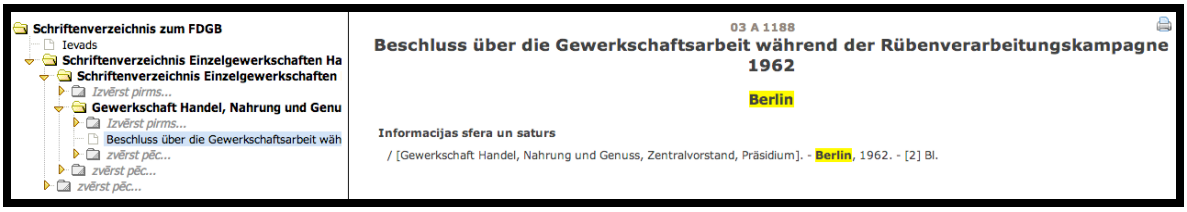

Figure 5. Mismatch between interface language and metadata language

Some websites offer metadata in different languages. One example is the State Hermitage Museum[6], which offers all its content in English and Russian. Figure 6 and 7 show an example of a search result, which is in Russian or in English depending on whether the search was started from the Russian or English interface. Unfortunately, you cannot switch between languages during your search. It is also not possible to search in Latin alphabet under the Cyrillic interface and vice versa. In practice this means that these are two separated instances for each language that are not connected. Truly multilingual search is not possible with this implementation.

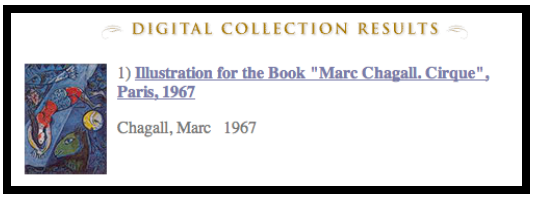

Figure 6. The State Hermitage Museum result display in English

[4] http://translate.google.com/

[5] http://www.archivesportaleurope.eu/Portal/

[6] http://www.hermitagemuseum.org/

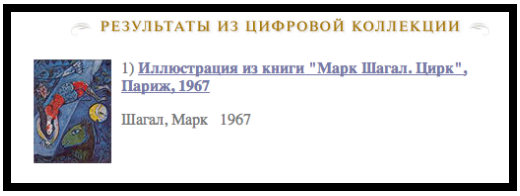

Figure 7. The State Hermitage Museum result display in Russian

The display of multilingual objects needs to take into account all the different dimensions of multilinguality explained in chapter 2. Not only is the provision of this data a challenge but also the visual design and usability. The International Children's Digital Library is providing access across languages to its objects. Figure 8 gives an example of an object that is originally published in German. The whole book was digitized and can be read on screen in German and several other languages that were added by volunteer translators. The display summary has its own language drop-down menu that translates the field values in the preferred languages, in this case Spanish.

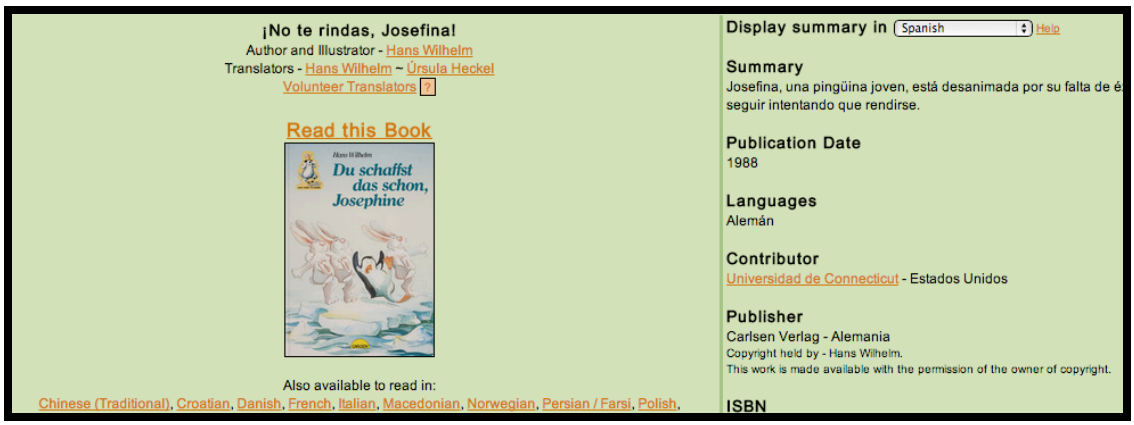

Figure 8. ICDL – German book with Spanish summary

This is an example of a multilingually rich offering for users with different language backgrounds. In terms of display the summary can be changed via the commonly accepted drop-down menu. The different language versions of the book can be accessed via links. Here the special display is that each translation was embedded in the digitized page as if the book would be printed in these languages (figure 9).

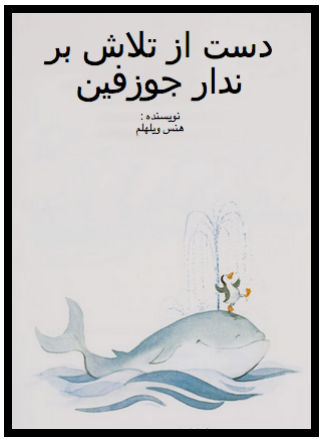

Figure 9. ICDL – translated book cover

4.3 Multilingual Search

Multilingual information retrieval is rarely implemented in the cultural heritage domain. This might be due to the technical implications as cross-lingual information retrieval is requiring query or object/metadata translation. None of the websites had query translation implemented. Nevertheless, more and more sites enable search in different languages through multilingual controlled vocabularies. Especially, in fielded location search, vocabularies such as GeoNames[7] are used to enable cross-lingual retrieval. Another approach is to use the Google Maps API that comes with cross-lingual retrieval for geographic names. Figure 10 shows the MapRank Search of the David Rumsey Map Collection[8]. Based on the locale of the browser entering the site the search result is shown in in the according language. However, it retrieves the correct geographic location no matter in which language the query was expressed. In the example here the query is for the German town 'Köln'. The map search retrieves the correct entry although the result is in English with a different language version of this particular city.

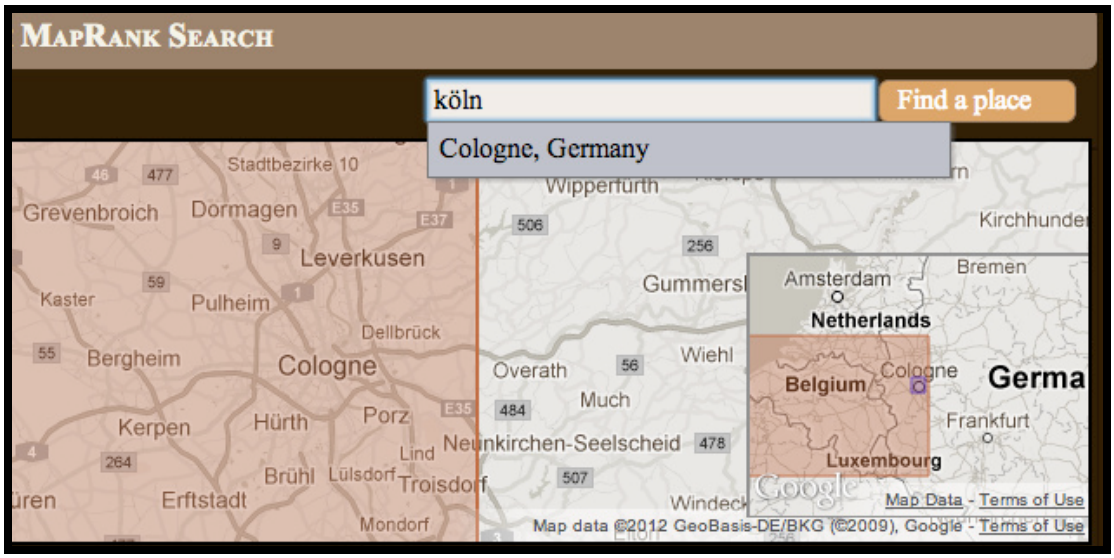

Figure 10. MapRank search of the David Rumsey Map Collection

More and more institutions try to offer alternative access points to their material beyond the search box. Searching an information system with a query requires an information need. Many institutions want to offer different access that is closer to the experience people have with cultural heritage. The International Children's Library[9] provides several access points that are targeted towards their domain: books from the childhood. Therefore, they offer a wide range of search features that are very helpful re-retrieving a book that was read many decades ago. User cannot remember authors or titles but rather the color, shape or main character of the book. The Library managed to offer these features enabling discovery across languages. Figure 11 shows a screenshot of the library and their different buttons which act as facets to drill down the number of results. It offers the facet to minimize search result based on the color of the cover and on the characteristics of the main character. Furthermore, it has uncommon facets such as 'long books' or 'make believe books'.

[7] http://www.geonames.org/

[8] http://rumsey.mapranksearch.com/

[9] http://en.childrenslibrary.org/

Figure 11. ICDL – alternative facets for children´s literature

4.4 Multilingual Browsing and Exploring

Multilingual browsing is rarely implemented in the sample cultural heritage websites. Nevertheless, more and more browsing features are not based on textual content anymore but rather on spatial and time browsing which can be understood across languages. Many websites use APIs of map providers to allow users different access points to the material. Map and timeline browsing allow viewing the collection from a different perspective. Furthermore, it supports the user to find unknown objects, which are not only based on semantic similarities, but on spatial closeness and chronological relations.

How well these features work to discover content in unknown languages depends on the underlying metadata quality and on the user interface like following common conventions on timeline usage.

Map browsing

Browsing maps is a very convenient way to enable the user to discover items based on their geographic location or their connection to a certain geographic location. It is essential to make this connection visible to the user: Are items shown on the map located in this place or do they carry information about this place?

Figure 12 shows the map-browsing interface of HistoryPin[10]. Here it is possible to navigate the content without understanding the language of the metadata as the user can browse to the country he is interested in: pins indicate objects connected to this particular location.

[10] http://www.historypin.com/

Figure 12. Map browsing interface showing objects connected to certain places

Timeline Browsing

Timeline browsing is anther way to bridge language gaps and provides users with different access points to content. Figure 13 presents a picture of the combined map and timeline entry point of the World Digital Library[11]. Results can be narrowed down on the map with a slider on a timeline.

Figure 13. World Digital Library offers timeline browsing to narrow down results

Search by Color, Shape or Layout

In general, more and more different access points are explored. They are mainly targeted on exploiting textual information, which can be found in metadata of the digital objects. Recently, characteristics of the digital objects itself like shape and colour are used to find connections

[11] http://www.wdl.org/

between objects and enable users to explore collection in different ways across languages. The Rijksmuseum[12] offers refinement of search results by colour (figure 14).

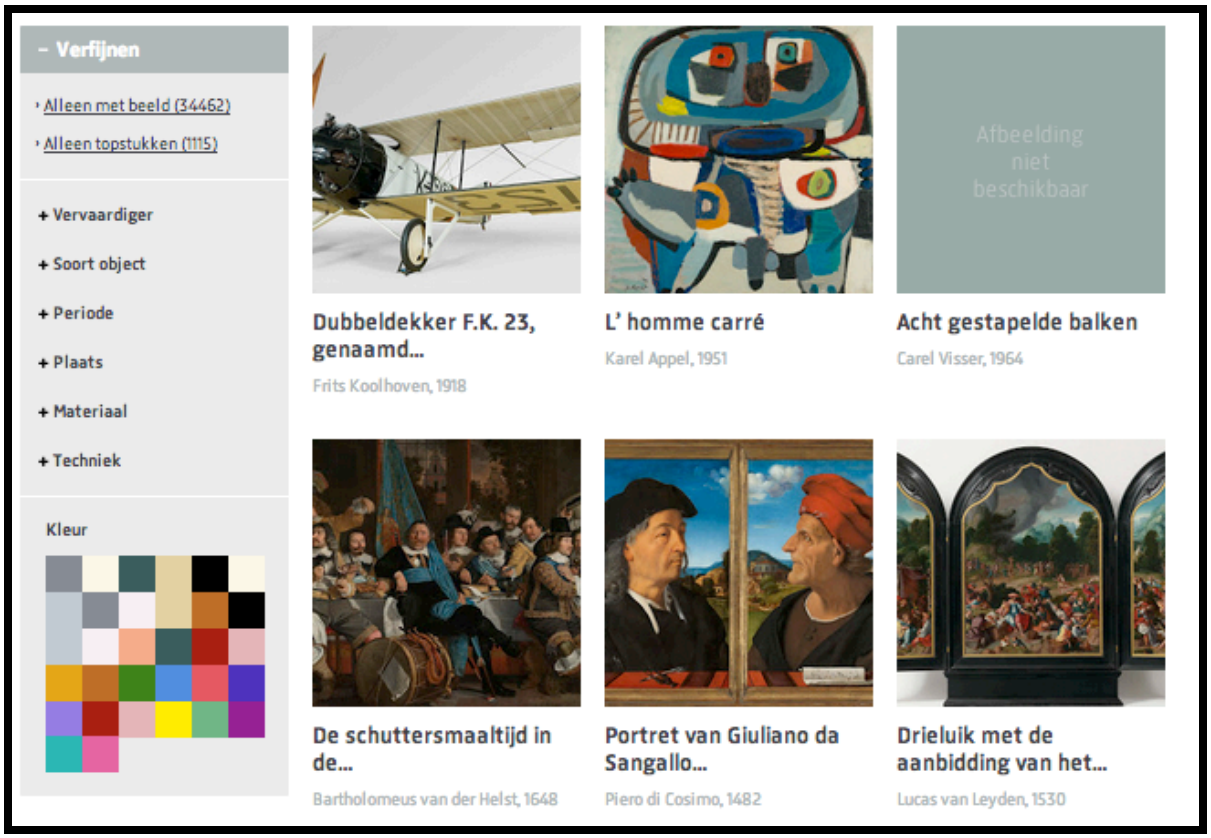

Figure 14. Rijksmuseum allows the user to refine search results by color

The State Hermitage Museum[13] provides layout search. The user can specify a layout with shapes and colors and the system will retrieve objects that are arranged in a similar way (Figure 15).

[12] http://www.rijksmuseum.nl/

[13] http://www.hermitagemuseum.org/

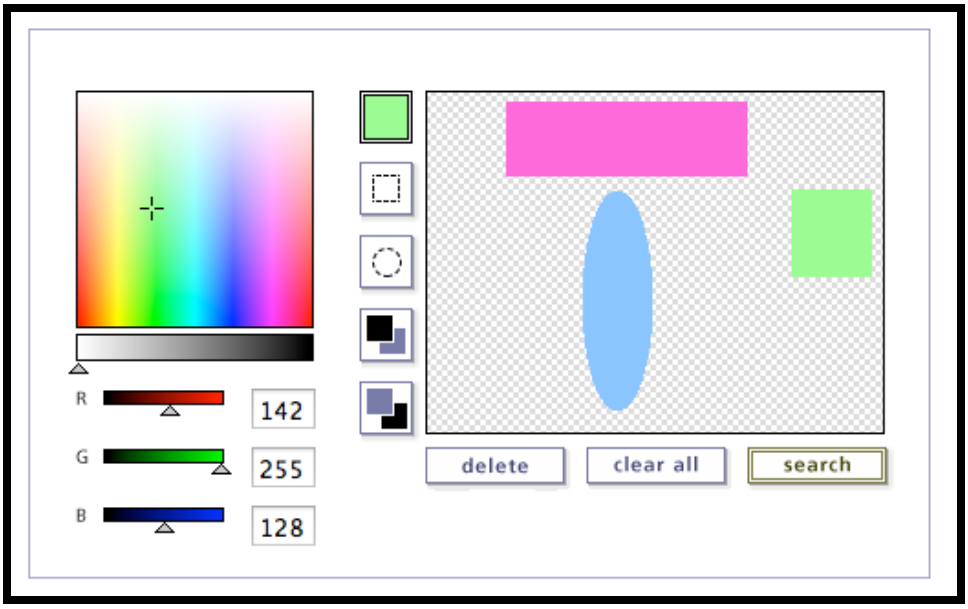

Figure 15. State Hermitage Museum offers layout search

4.5 Collaborative Features

Only a few websites in the cultural heritage domain offer collaborative features for users. Reasons for this are manifold, on the one side there is the fear that user input might result in poor quality content, on the other side there is the technical implementation of such a system. They are complex requiring user rights management implementations. Another pitfall is to set the right incentives for users to participate. Many websites in the cultural heritage domain offer great features but are missing a solid user base. In the following paragraph, implemented collaborative features, namely social tagging and collaborative translations will be elaborated.

Social Tagging

The most common implemented feature is social tagging, generally accepted to add value; implemented in the right way it can enrich the metadata. However, most of these social tagging features are not designed to aggregate multilingual tags or annotations; they are targeted for monolingual use. A strategic implementation of social tagging guiding the user through the workflow ensures a high quality of the tags. The project Your Paintings[14] motivates people to tag paintings from museums from the UK. In a step-by-step process, users are guided through the process making sure tags are added in the right category, named person, event or things. This way a semantic layer is added to the tags, which makes them even more useful (figure 16).

[14] http://tagger.thepcf.org.uk/

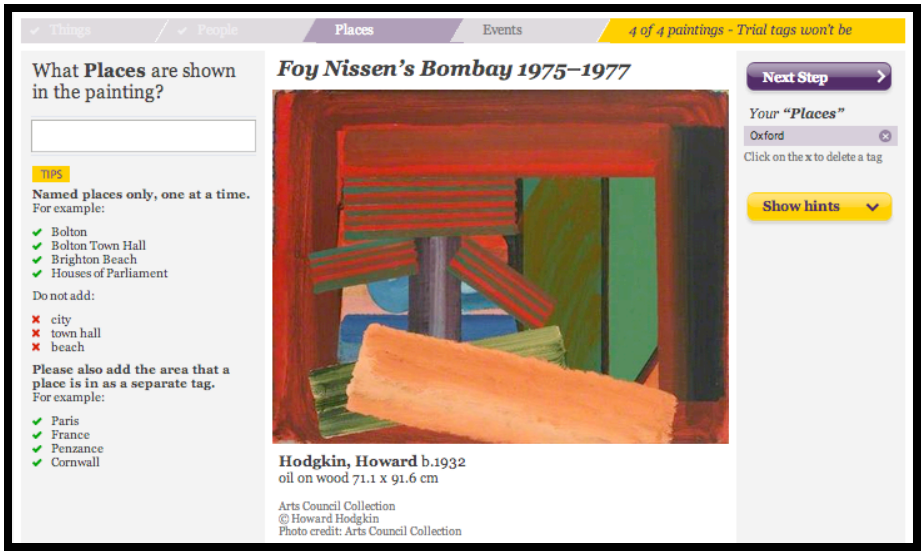

Figure 16. Guided tagging process resulting into semantically enriched tags

Dictionaries and controlled vocabularies reduce the risk of misspellings and help the user to disambiguate homonyms and named entities. The Steve tagger project is the only website which offers the user to specify the language of the tag (figure 17). Such a feature can be used to improve retrieval across languages.

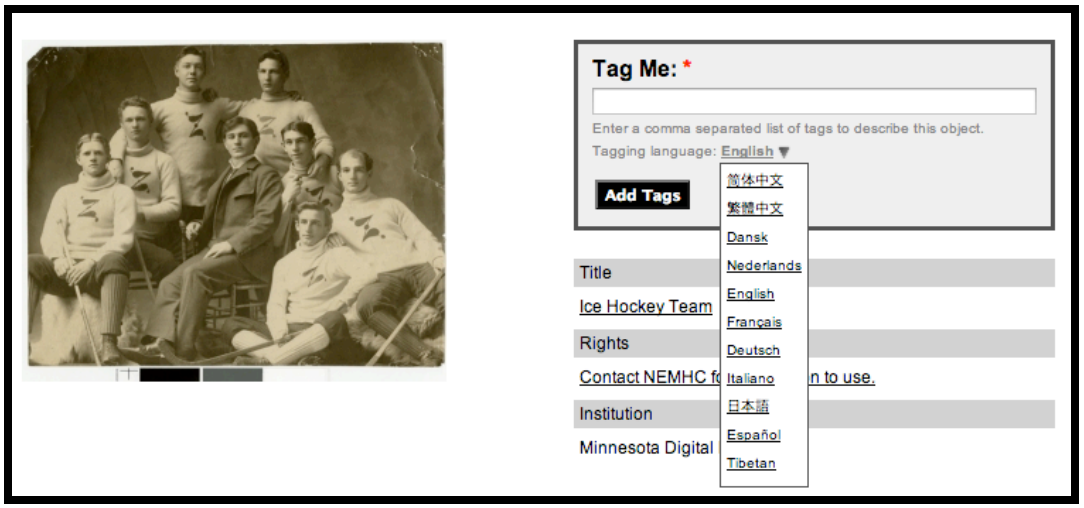

Figure 17. Steve Tagger Project – drop-down menu to specify the language of tags

Collaborative Translation

Active calls to participate in enriching metadata and helping adding more information to digital objects are rare and in the most cases they are dealing with monolingual content. Most of these initiatives fall into the category of citizen science where users are invited to collaborate on certain tasks such as transcribing scanned manuscripts, e.g. Transcribe Bentham[15].

Collaborative translation features require complex systems to manage the user input and provide processes, which ensure quality. To a large part cultural heritage websites manage translation loads they want to outsource to the public by recruiting volunteers. In this case, the community does not oversee the quality and process of the translation; this is rather done by individuals who are managed by the offering institution. For example, the ICDL recruits

[15] http://blogs.ucl.ac.uk/transcribe-bentham/

volunteers who translate the books, metadata and the web interface in other languages (see: http://en.childrenslibrary.org/contribute/translate.shtml). Quality control and the responsibility for correcting the translation stay with the institution. It is obvious that this solution requires a high level of maintenance and the process is not self-sustainable. On the other side, handing translation completely to the community requires a community strategy, an engaging and usable web interface and community managers who enforce rules and guide the process.

4.6 Summary

This survey of cultural heritage website showed that many sites deal with multilingual issues such as users from different countries and objects in different languages. Nevertheless, multilingual access is mainly limited to offering the interface in several languages. Only in rare cases this is extended to the metadata of the objects.

Multilingual search and browsing is applied rarely, only if the metadata and the keywords are in several languages. None of the sites offered query translation to support cross-lingual information retrieval. However, many sites explored the possibilities of browsing features, which are not based on textual indicators such a titles and metadata field values. Map browsing and timeline browsing are implemented often and offer access to content overcoming language barrier.

In terms of enriching metadata with user-driven data, many websites rely on social tagging. Implementations of effective collaborative social tagging features are rare but positive examples described above show that with a right strategy the process could be guided to receive the highest quality of tags.

For Europeana, this means that in terms of multilingual access it can be a trailblazer guiding the direction for future developments.

5. Multilingual Interactions in Europeana

In this chapter, the multilingual access points and strategies of Europeana are described. First steps of implementations were taken. This section analyses the status quo in multilingual access to Europeana's content. It gives suggestions for improving existing features to make them truly multilingual acknowledging that some of them might be hard to implement.

5.1 Use Case Europeana – Research in Multilingual Access

In the past, Europeana functionalities were evaluated by end-users and experts investigating user behavior and expectations through surveys (IRN-Research, 2009), log file analysis (Clark et al., 2011), usability studies with focus groups (Dobreva and Chowdhury, 2010) and workshops (Ferro and Petras, 2009). In 2009, a survey focusing on multilingual access to Europeana was conducted (Agosti et al., 2009) including 7 topics:

- User profile including native & other languages and digital library use
- Multilingual content interaction
- Multilingual user interface
- Information access and retrieval
- Multilingual information retrieval
- Multilingual query formulation & expansion
- Multilingual results presentation

The survey was introduced during the TrebleCLEF Summer School on Multilingual Information Access (http://www.trebleclef.eu/summerschool.php). The 25 participants were frequent web and/or digital library users and came from 13 different countries. The majority of users (80%) were willing to control the query translation process. No clear preference was found regarding the multilingual result representation. However, the results of an online survey determined that the most popular result refinement options are the language and country facets (IRN Research 2009).

A second survey included questions concerning the users' native language and language skills and was provided in six languages. On average, respondents had language skills in at least 1.5 other languages and 71% of all non-English native speakers could access and interact with websites in English (IRN Research 2011).

In general, users feel comfortable accessing the portal and scanning results in their native language or in English. A "significant language barrier was perceived" when users had to deal with content in unknown languages (Dobreva and Chowdhury, 2010). Dobreva et al. also found out that a stronger need for more content in native languages as well as result translation options exists.

Within the EuropeanaConnect project http log files from October 2009 to September 2011 were analyzed with regard to general access statistics (Clark et al., 2011). Most users were coming from France (16%), followed by Germany (14%), USA (10%) and Poland as well as Spain (each 7%). A high preference from users for collections from their own countries could be observed. Due to the fact that Europeana does not offer cross-language search, this might be a result of queries in a particular language only matching metadata from the appropriate national content provider.

In this section, the different multilingual dimensions in Europeana are listed. Furthermore, challenges are outlined with regard to the implementation and interaction design of the multilingual feature and how they can be addressed.

The implemented features were described in the deliverable on the outline of the functional specification for Europeana (Dekkers et al., 2009). The main focus was put on the implementation of a multilingual interface, browsing capabilities, multilingual search in its different facets and result translation. Most of the specifications were developed but not all of them implemented in the Europeana production system. For example, EuropeanaConnect developed a query translation module that can detect the language of a query and translate it into 10 different European languages.

Furthermore, a report written in EuropeanaConnect describes the different access strategies in Europeana but was not focused on the user interaction and usability concerns regarding their implementation (Petras, 2011). In this section, it will be elaborated on.

5.2 Multilingual Display in Europeana

Accessing the Different Language Versions of Europeana

There are several ways for users to access Europeana in their native language:

1. By default, the English interface is shown and the user can switch to his desired language by choosing it from a drop-down menu (figure 18).
2. Once the user switched to his preferred language, a cookie is set which directs him to his selected interface on the next visit.
3. Until recently it was also possible to access Europeana via a customized link that carries the language parameter. This link was visible in the browser bar when switching the interface language to a different one than English. Figure 21 shows a picture of this URL ranking in Google. A German user can directly choose to access Europeana with a German interface.

Interface Language Change

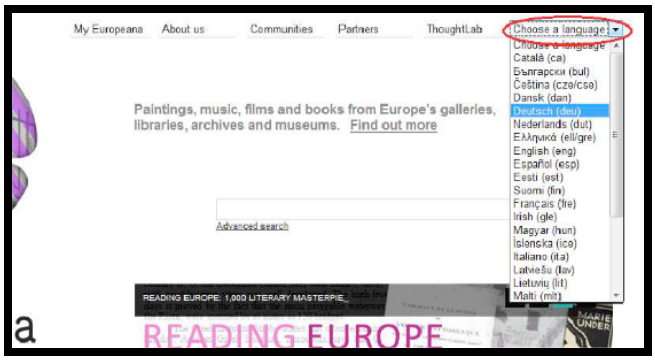

Figure 18. Europeana – Interface language change via drop-down Menu

Europeana offers a multilingual interface, which can be switched via drop-down menu. So far, the interface is translated and can be shown in 31 different European languages. Once the user chooses a language, a cookie is set, so that on the next visit, Europeana is delivered in the right language as long as cookies are enabled and were not deleted in the meantime.

➔ **Challenge: Only translation of static content.**

A lot of content in Europeana is pulled dynamically; unfortunately this content is not part of the language-skin.

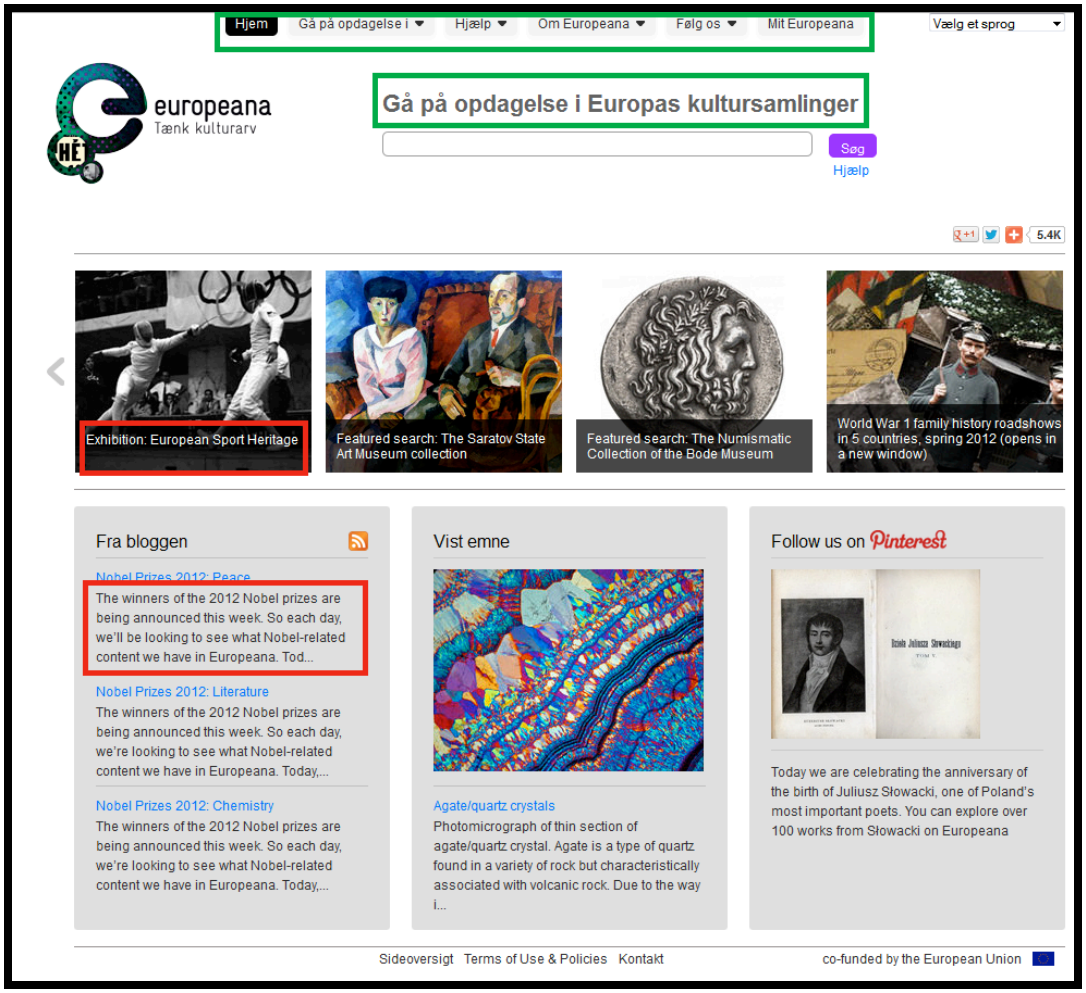

Figure 19. Europeana – inconsistent Danish interface with English text (marked red)

Furthermore, it is not clearly labelled that the interface language change does not change the language of the collections.

Suggestion
The dynamically pulled in content should be also translated to the language the user prefers. Otherwise the page is a mix-up of different languages. As the homepage mainly features dynamic content, most of the text is in English although the user switched it to Danish as figure 19 shows.

➔ **Challenge: Users might think that the interface language is connected to the language of the search.**

It might not be clear to all users that the interface language drop-down menu only changes the language of the interface. Some users appear to thing that this also affects the search and collections searched in. It should be visible that the drop-down menu only affects the language of the static content and not the search or browsing functionalities of the system.

Suggestion
An explanation of the affects of the drop-down menu could be added in form of a questions mark explaining the effects of the language change.

Referrer links with language parameters

➔ **Challenge: Generic URL in browser address bar without language parameters.**

For now, it is impossible to see in the address bar of the browser which language version you are on. The parameter gets overwritten and the user sees a generic URL (figure 20). The problem is that specific language version of Europeana cannot be linked to, as the URL with the right parameters is not shown. So every new visitor when following a referrer link from a search engine or third-party sites is likely to end up in the default English interface.

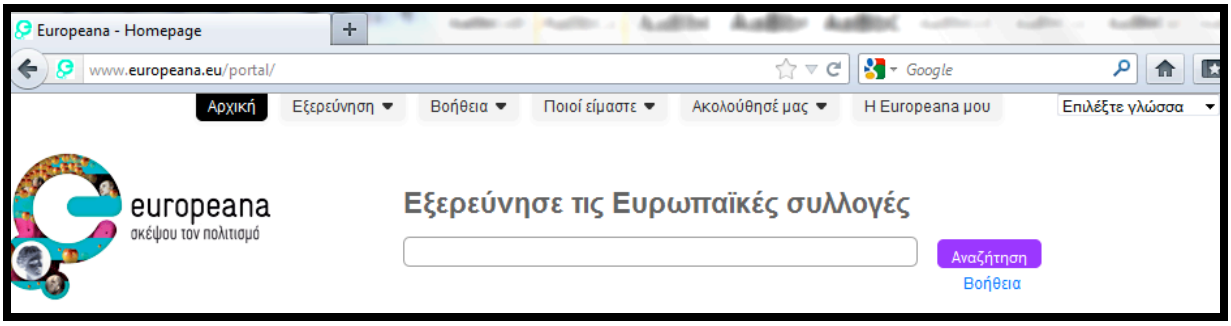

Figure 20. Europeana - Generic URL in address bar although the Greek interface is selected

Furthermore, these URLs with a language parameter seem to exist and if known can be used. They are in the form of www.europeana.eu/portal/?lang=language
Figure 21 shows a screenshot from 2011 where you can see that Google indexed them and offered them as a localized version to users searching in a local Google version. That means the French Europeana was offered in the French Google. Now these URLs do not rank anymore. And it seems to be an issue of setting canonical URLs without specifying the different language versions.

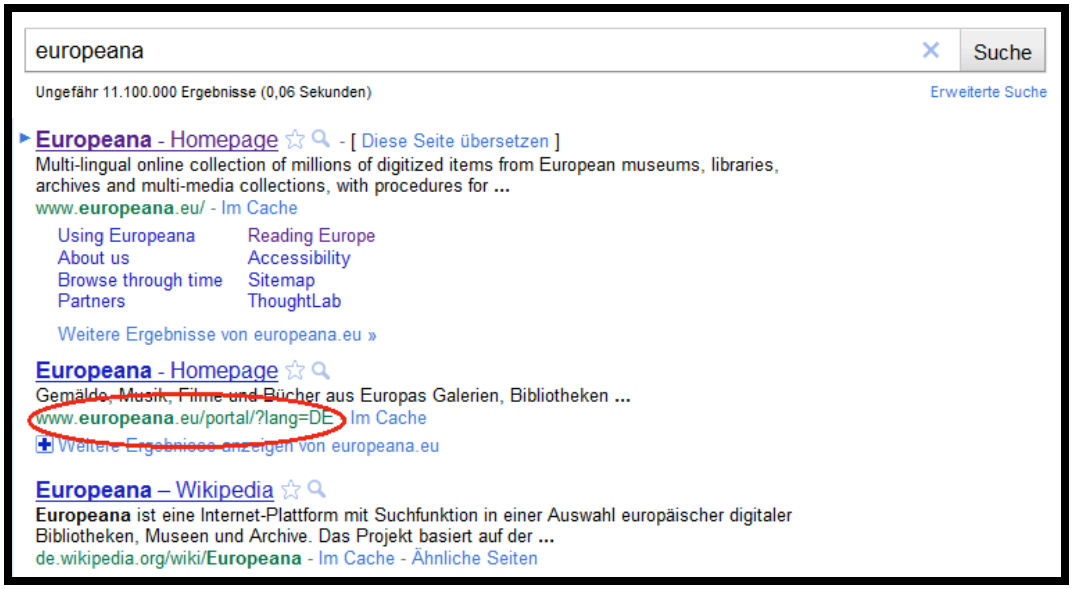

Figure 21: Language specific links indexed in Google

Suggestion
Add language parameters to URL to enable users to access the appropriate language version via external links.

5.3 Multilingual Search

For supporting search across languages and the expression of information needs in a query, Europeana provides a suggestion or autocomplete feature. It supports users in the query formulation process and helps to avoid misspellings (Hearst, 2009).

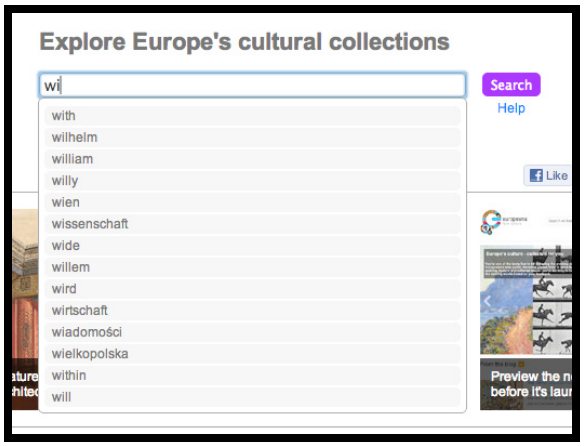

Figure 22. Europeana – Query suggestion or auto-completion

➔ **Challenge: Users might be overwhelmed by unknown terms in languages they do not understand (see auto completion in figure 22).**

Suggestion
Offer the user either only suggestion in one language (e.g. depending on interface language) or visually separate the suggestion coming from different languages. This could happen through colour coding for every language.

Object Translation

Result or object translation is currently offered by an external translation service provided by Microsoft. Previous studies have shown that users are satisfied with metadata translation and do not require full text translation in order to assess results (Oard et al., 2004).

As shown in figure 23 users can select their preferred language via a drop down menu and translate metadata information including the object description. The object title as well as the metadata field names are not translated but remain in the source or selected interface language. At any time users can go back to the original representation. Translated objects are not stored in the user profile but only the original version is saved.

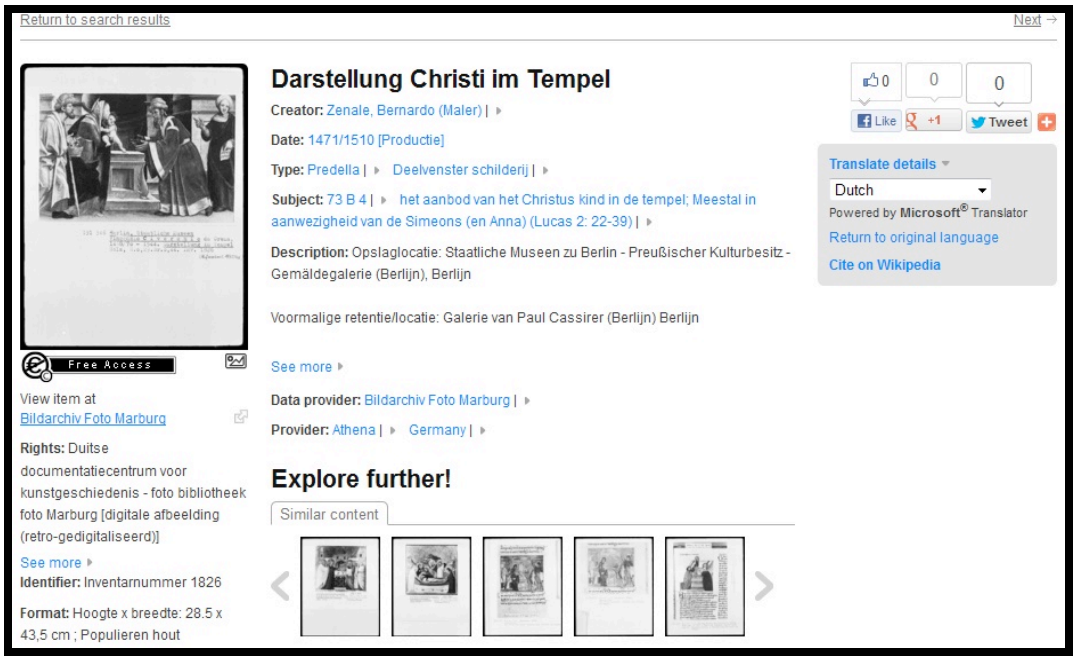

Figure 23. Result translation from original language German to Dutch

➔ **Challenge: Partly translation may lead to confusions since it is not obvious to the user why some fields or descriptions are not translated.**

Suggestion
Object translation should be consistent and also include the translation of metadata field names. Although this is normally achieved through the interface language change it would be easier for users if only one interaction is needed here for the translation process. User might have the possibility to store translated versions of an object in their user profile.

5.4 Multilingual Browsing

Multilingual browsing capabilities are essential for digital libraries to enable the user to understand extend and scope of a given collection and support serendipity and discovery of unknown cultural material in languages which are not understood by the user.

Europeana Exhibitions

Europeana curates virtual exhibitions, provided by different European institutions, around thematic themes. Most of these exhibitions are offered in different languages (figure 24).

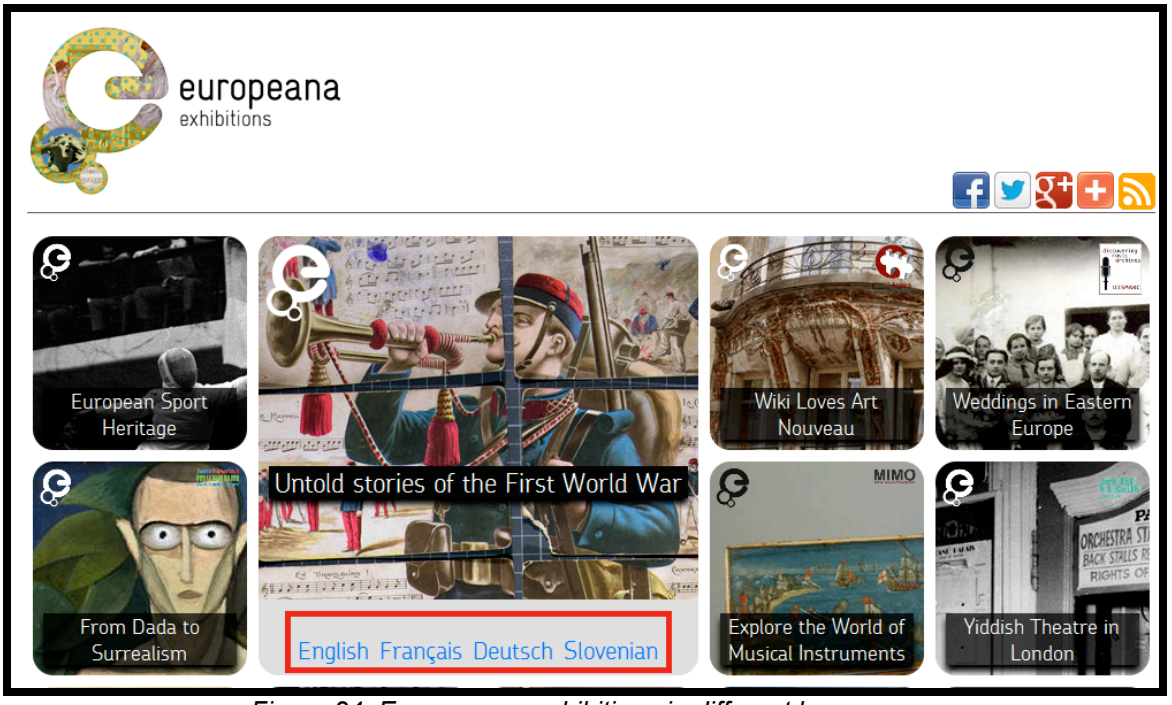

Figure 24. Europeana - exhibitions in different languages

➔ **Challenge: The language of exhibition is only visible when expanding the view of the thumbnail.**

Suggestion
The query plug-in Isotope[16] which is used here also offers the possibility to add filters and sorting functionality. Language sorting and filtering eases access to language specific exhibitions and allows the user to get a quick overview about exhibitions in his language.

Europeana Timeline and Map display

There are two language-agnostic features that enable users to see results for their queries on a timeline or a map. Theoretically this could enable browsing across language but the features are only special result displays that require a query. Results are than mapped on the geographic location the objects is provided from and mapped on a timeline according to the date specified in the metadata. Both features would need a complete overhaul to be more self-explanatory and user-friendlier.

Result Filtering by Language and Country

Europeana offers several facets to filter retrieved results. The language facet allows users to refine objects according to the metadata language. Note that the language of a metadata description does not necessarily correlate with the actual language of an object. A document provided by a German library would appear as a German object even though it might be written in several languages. Figure 25 shows a result page refined by the German language facet. Once this facet has been chosen, only countries with German as official language are displayed (in this case Germany and Austria).

The language facet “multilingue” includes objects from content providers that aggregate several institutions such as The European Library.

[16] http://isotope.metafizzy.co/

Through the country facets users can refine their search to objects provided by institutions from one or more countries. Again, the origin of the content providers does not always indicate the language or origin of a particular object. Equivalent to the language facet "multilingue", a country facet "Europe" exist.

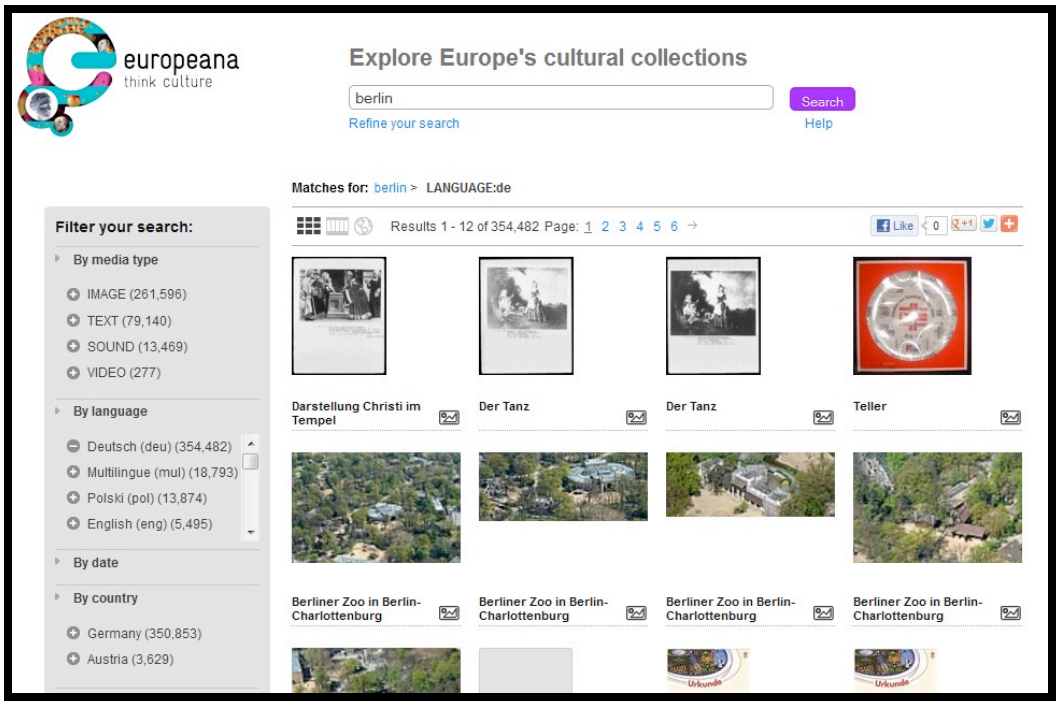

Figure 25. Europeana – Result set filtered by language: German

➔ **Challenge: Especially for textual documents, users might be confused or disappointed if documents in languages are presented different to the ones they chose.**

Suggestion
It should be clear and transparent to users that the language and country facet are associated to the origin of the content provider and d not necessarily represent the object language.

5.5 Multilingual Semantic Enrichments

Europeana enriches their content with multilingual vocabulary to enhance metadata multilingually and enable retrieval across languages (Olensky et al., 2012)[17]. The semantic enrichments were done with vocabularies specified in table 3. For each field that was enriched, a specialised vocabulary was used.

[17] Parts of this subsection were taken from this publication at MTSR 2012, which emphasized the need of a multilingual and semantic enrichment strategy for Europeana. The publication was part of the research done in task 7.3 and 7.4 of the Europeana v2.0 project.

Vocabulary	Website	Type	Enriched metadata fields
GEMET Thesaurus	http://www.eionet.europa.eu/gemet/	Concept	dc:subject dc:type dcterms:alternative
DBpedia	http://dbpedia.org/About	Agent	dc:contributor dc:creator
Semium Time Ontology	http://semium.org/time.html	Period	dc:date dc:coverage dcterms:temporal
GeoNames	http://www.geonames.org	Place	dc:coverage dcterms:spatial

Table 3: Controlled vocabularies and structured datasets used to enrich Europeana's metadata fields

With regards to multilinguality, these enrichments are of utmost importance. For example, enriching subject metadata fields with all translation equivalents of a concept enables the user to retrieve documents, which are not written in the language of his query. Figure 26 shows an example of the query “cheval” (fr. horse) which retrieves a Russian object, which has the Russian term for horse in its title ‘Лошадь’. This term now got enriched with translation from the GEMET thesaurus. The associated tags can be found in the right side bar under ‘Auto-generated tags’.

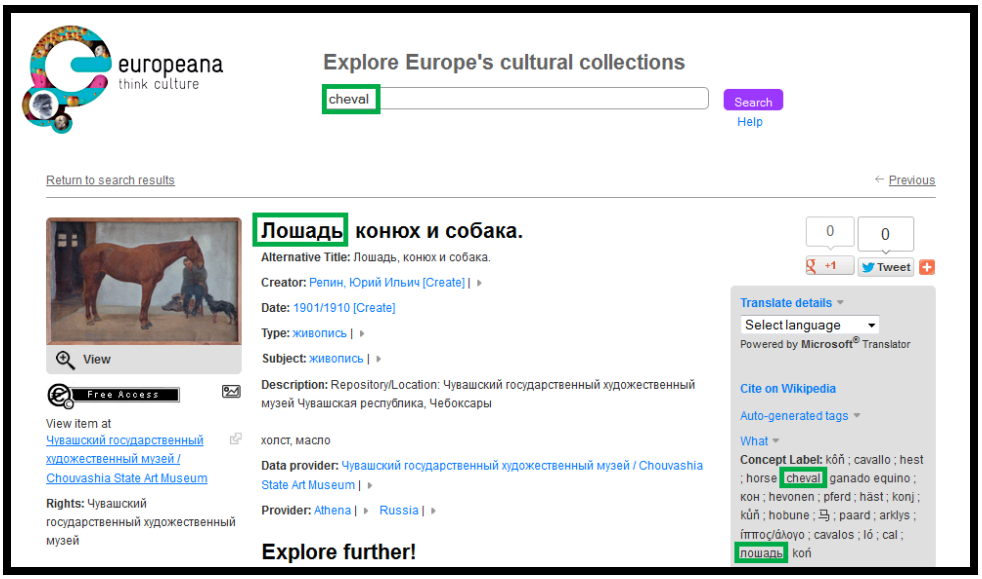

Figure 26. Result for Query “cheval” retrieved based on multilingual enrichment of concept term

➔ **Challenge: Cross-lingual ambiguities might lead to confusing enrichments and poor user experience.**

One example of this shortcoming in Europeana’s enrichments is a search for ‘poison’ in the collections of Swiss Institutions that will return photographs from India and Indian movie covers. A correlation between the user’s query and the retrieved object is completely missing, resulting in a poor user experience. The reason for this error-prone retrieval result is the automatic enrichment of the term ‘Inde’ (fr. India), which was a concept label for the

retrieved French objects. In Latvian 'Inde' means poison. This was used as a basis for enriching all concept terms 'Inde' with *poison* and its language equivalents.

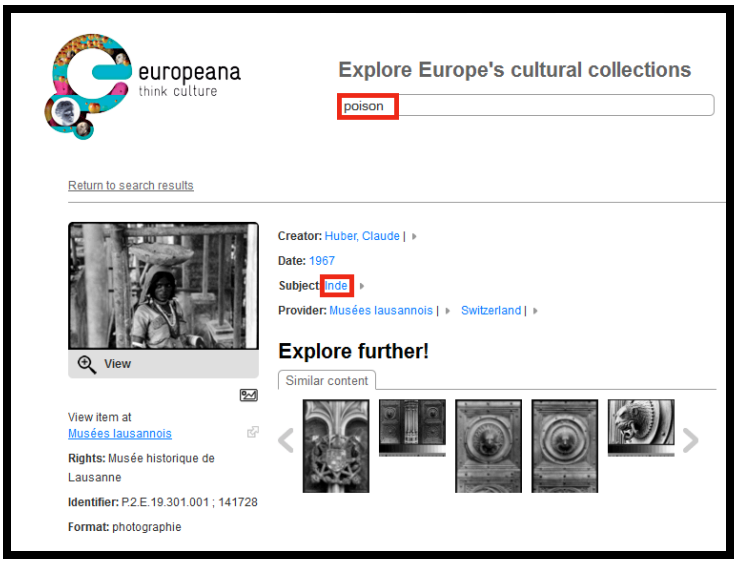

Figure 27. Irrelevant result for query "poison"

Suggestion
To avoid cross-lingual ambiguities (words which occur in several languages but have completely different meanings), metadata records and enrichment terms need to have the same language. If the language of the metadata is not known, the language derived from the provider country can be sufficient.

Another area where enrichments have an effect on multilingual retrieval is when enrichments emphasize keywords in small languages pushing the retrieval performance of enriched objects. One example is an object, which depicts a piece of traditional clothing for women[18]. Amongst the many keywords, which describe this object, the Romanian term for *woman (femeie)* was chosen. If a user now looks for the term women in his language trimming down the result set to the Romanian results, he will get the enriched object, although many more Romanian objects would be more relevant to the query (because they actually depict a women). Although the particular object gained more visibility it is retrieved for a query to which it is only broadly relevant.

➔ **Challenge: Enriched term put a lot of weight to the object they are expressing. In some cases the chosen term for enrichment is not the most suitable surrogate for the object in different languages.**

Suggestion
Define matching rules for digital object with several potential enrichment terms. Some are so broad that they are not suitable for enrichment. The enrichment rule should determine under which conditions enrichment should occur.

18 http://www.europeana.eu/portal/record/05812/7BDA67A91EABA04D4CDDDE01F400B34FAB08A9A1.html

Europeana Ingestion process[19]

Multilingual metadata can be ingested in Europeana if xml:lang tags in all appropriate metadata elements are specified, i.e all those elements that have a text string as a value. So far, there is no multilingual functionality in Europeana, which allows leveraging this tag, but it is expected to be developed in the near future. Metadata records in different languages for the same object cannot be linked in the portal right now and would lead to redundant objects that cannot be connected. They way to go here for providers is to indicate the language for each value in the metadata field, so it can be displayed.

➔ **Challenge: Records with metadata values in different languages cannot be displayed (Figure 28).**

Figure 28. Keywords in several languages in metadata fields cannot be displayed in a user-friendly manner.

Suggestion
Indicate the different languages in fields and visually separate them from each other.

Only one instance of the dc:title can be displayed for now. To ensure that translated titles are getting displayed, different language versions of titles should be put into dcterms:alternative.

[19] Based on work in task 7.4, the FAQ for providers were updated to reflect the status on ingesting multilingual data. Some of the content developed for these FAQs was used in this partagraph. The FAQs for providers can be found here: http://pro.europeana.eu/web/guest/providers-faq

6. Use Case for User-Assisted Translation in Europeana

User-assisted translation either makes use of indirect user input such as query logs or directly involves the user into the translation process. Different approaches to leverage user data have been presented and discussed in section 3.2. It is still an open issue how the quality of user-generated input should be controlled and measured. Interactive systems need to support and encourage the user to participate in the search process. Simultaneously, the workflow should not require too much effort from the user's side and required clicks need to be minimized.

User-assisted translation is a multi-level process that includes several steps where users can interact with the system:

- Determination of source language
- Determination of target language(s)
- Translation selection
- Result translation and
- Object translation

Figure 29 summarizes the 5 steps that need to be taken into account when implementing user-assisted translation functionalities.

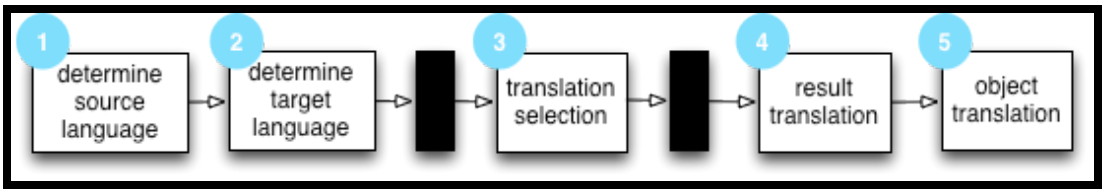

Figure 29. Interaction points for user-assisted translation

(Peters et al., 2012) provide an overview of important aspects regarding the implementation of query formulation and translation to multilingual information systems.

Interaction Point	Description
Detection of source language	Systems can either automatically detect or ask the user to specify the source language.
Correlation between source and interface language	Systems correlate the preferred interface language to the source language. In some cases users might not want to search in different languages than the interface language
Selection of target language(s)	Users might be interested in results provided in different languages. At the same time users might want to languages they are not familiar with.
Named entity detection	In order to reduce wrong translations user could indicate named entities, which will should be handled separately.
Translation selection / user-generated translations	Users might want to select the most appropriate translation candidates suggested or even want to add another translation to the list. User-created dictionaries can be used to disambiguate and improve search results

Table 4. Aspects of query formulation and translation (Peters et al., 2012, p. 99 - 100)

6.1 Determination of Source Language

In the beginning of a search, the user needs to determine the query language for further processing steps. Different to systems that need to identify the query language automatically a system that allows users to select the source language avoids problems inherent with query language detection. Figure 30 presents a search for “Shakespeare Biografie” with two options to indicate the query language. Depending on the number of languages supported by a system, either a drop-down menu or a static list of languages is provided.

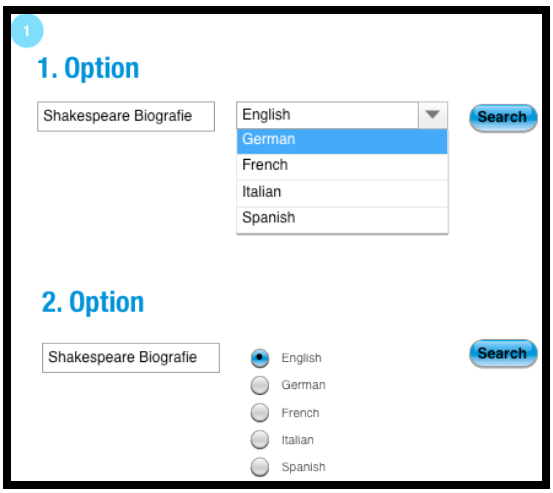

Figure 30. Determination of source language

Once a user determined the query language the system should remember the preferred setting.

6.2 Determination of target language(s)

Similar to the source language determination, the identification of the target language(s) is an important step for the query language process by limiting the possible translation pairs. Especially for language independent named entities like “Shakespeare” it needs to be determined which target language(s) a user prefers. It should be the goal to minimize user efforts by offering a “select all” option for all available languages.

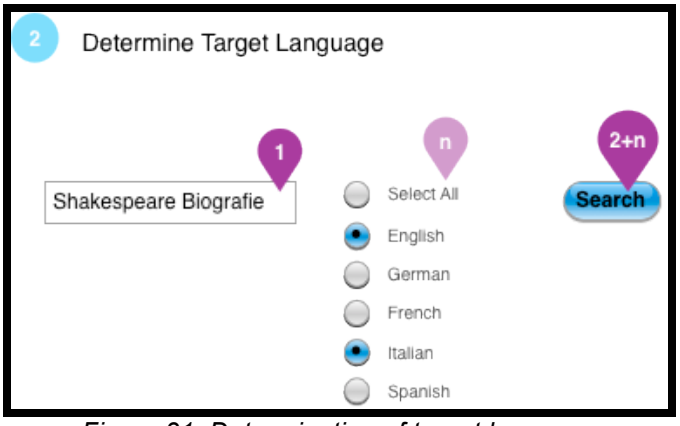

Figure 31. Determination of target language

6.3 Translation selection

The design and implementation of user-assisted query translation is not a trivial task. It needs to be determined how translation candidates are displayed, how users select appropriate translations and if users can correct or / and add additional translations. This is especially important for ambiguous terms. Figure 32 shows an example for translation

selection with translation candidates displayed separated by language. Another possibility would be a multilingual ranked list, displaying the most relevant translations first. For systems that support a high number of languages a high number of translation candidates appear and it is especially important to reduce non-relevant terms or suggestions. In this case the user can remove translation candidates and add own translations to the provided list.

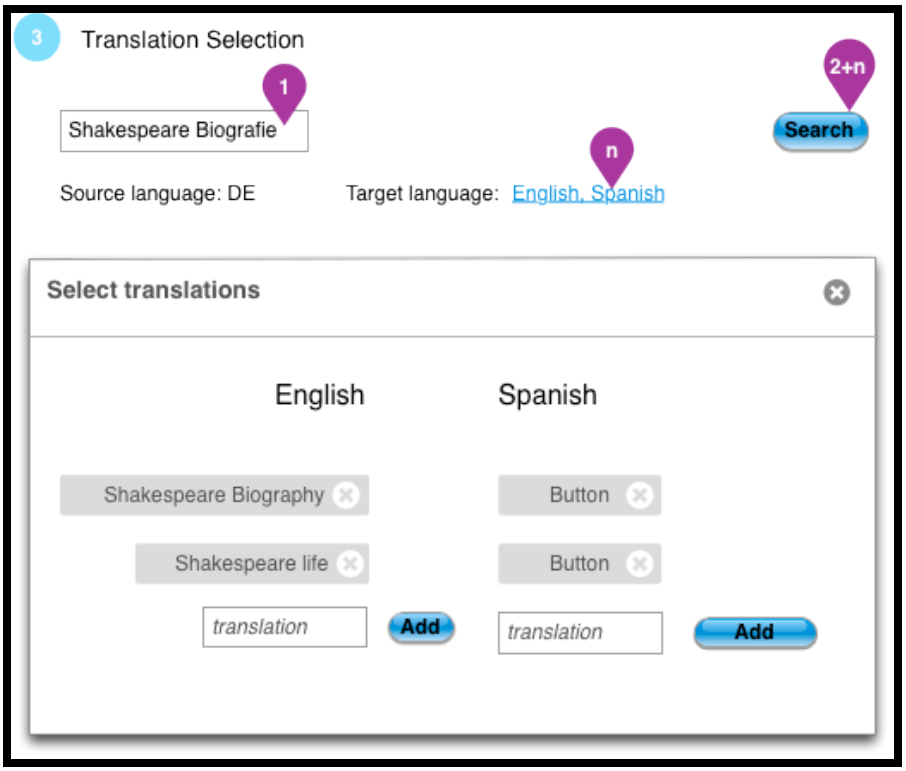

Figure 32. Selection and Adding of Translation Selection

6.4 Result Translation

The presentation of multilingual results can either be realized in one complete list or separated by languages. If the user selected more than one target language the system should offer language facets, which allow the user to refine results to one or more specific languages. Furthermore users can translate the result list into their native or preferred language.

Figure 33 demonstrates the result translation with a drop-down menu for language filtering and translation.

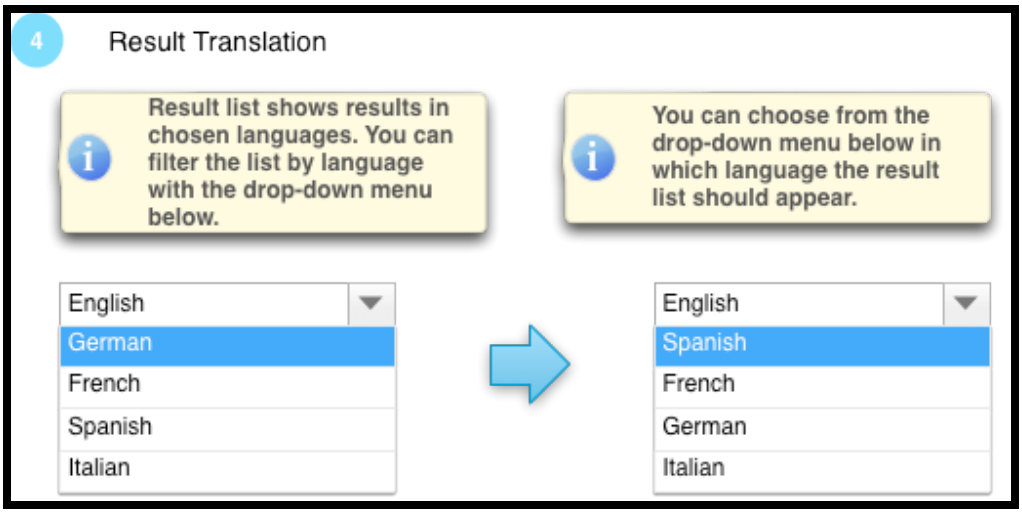

Figure 33. Result Translation

6.5 Object Translations

Object translation can either be achieved on the metadata level or on the object level. Especially for textual documents it needs to be determined what should be translated is translated and whether users are satisfied with metadata and abstract translation. Figure 34 demonstrates alternatives, the first option showing context translation and the second option providing translation options for metadata fields like title, creator or even a short description. This type of translation is used for non-textual media types like images, videos or sound files.

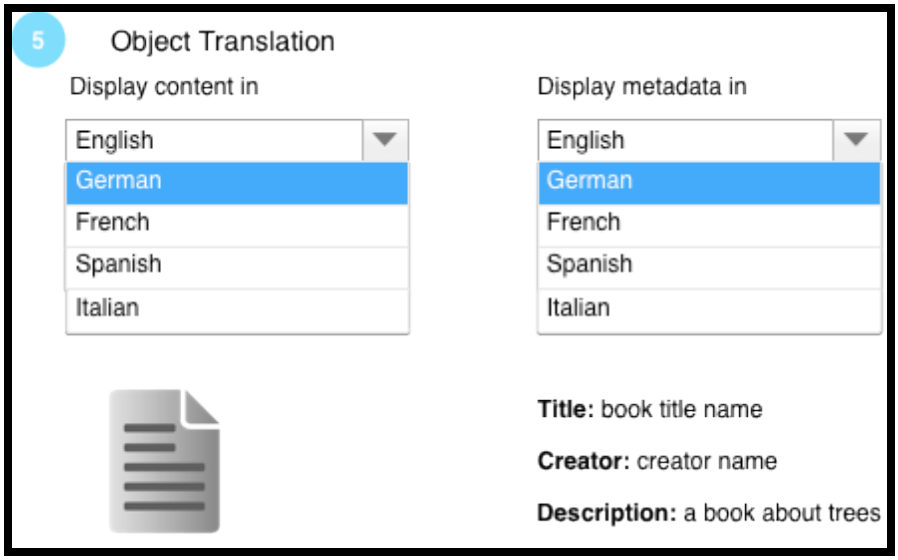

Table 34. Object Translation for Full Text or Meta Data Level

7. Multilingual Information Retrieval Evaluation for Europeana

In April of 2012, Europeana changed its ranking model from the vector-space based model used in the Europeana information retrieval system Solr to a customized BM25 model as proposed in the ASSETS project. For evaluation purposes, 85 multilingual queries taken from Europeana query logs were tested for performance using the new and old ranking model.

Europeana data and queries were also used for an evaluation lab (competition) at the 2012 conference of the Cross Language Evaluation Forum (CLEF)[20]. The Cultural Heritage in CLEF (CHiC)[21] pilot lab presented the results of different evaluation scenarios for the cultural heritage domain. The CHiC task scenarios were also used to evaluate the Europeana ranking for multilingual information retrieval.

The first part of this chapter describes the analyzed differences and the changes one could observe for the search experience between the old and the new Europeana ranking algorithms, whereas the second part briefly explains the CHiC evaluation scenarios and describes the outcome of the ranking algorithm evaluation.

7.1 Analyzing Europeana Ranking Algorithms

When Europeana changed its ranking model in April of 2012, the customized adaption of the text processing and relevance ranking based on work done in the ASSETS project was targeted towards improving the result listing for searches in Europeana. We compared the changes by extracting 85 queries from the Europeana query logs and running them against the "old" Europeana infrastructure and the "new" infrastructure.

Queries were selected and annotated according to the type of search such as named entity search (person, institution, geographical location, work title) or topical search (general topic, event). Previous research had show that most searches in cultural heritage systems are named entity searches and ambiguous in language (a query for "Mozart" is not necessarily asking for German-language material). The 85 queries that were selected for testing were chosen according to these criteria. For the CHiC evaluation campaign, which was held later, 50 of these queries were expanded with short comments suggesting an underlying information need, according to which retrieved documents were judged for relevance.

In the first part of the analysis, the retrieved documents were analyzed according to overlaps in the result sets and in the differences in ordering. A more detailed analysis looked at those queries where the result sets between the new and old infrastructure differed and the reasons for the discrepancies, which were mostly due to different query and document processing steps.

Result Set Analysis

Out of 85 queries, only 17 (roughly 20%) retrieved different document sets using the old and the new infrastructure. In almost all cases (with the exception of the query "quatremere"), the new ranking retrieved all documents found in the old ranking and then additional documents. The result lists for the query "quatremere" retrieved completely disjunct (although almost the

[20] http://clef2012.org/index.php?page=Pages/registrationForm.php

[21] http://www.promise-noe.eu/chic-2012/home

same in number) document lists and will be analyzed separately. Table 5 shows those query examples where the old and new ranking infrastructures retrieved different document sets.

Query	Overlap	Additional documents
quatremere*	0%	old: 21 / new: 23
dujardin karel	44.65%	207
friday 13th*	21.59%	138
jesse tree	85.83%	18
fishermen people	95.29%	4
monument globe	97.52%	3
jean jaques rousseau*	92.00%	2
luin	98.04%	1
paul colin	99.55%	1
pushkin	96.97%	1
quran	99.21%	1
samuel bochart	98.33%	1
silent film	99.59%	1
song medieval*	99.28%	1
swansea	99.75%	1
town crier	99.40%	1
unarmed	98.72%	1

*Table 5: List of queries with discrepancies in total results (document sets marked with * not caused by update of the Europeana collection)*

The analysis of the remaining 20% of the queries showed that there are two reasons for the differences in the result numbers.

The first was the dynamic nature of the Europeana collection, which added new documents during the time frame the different ranking algorithms were tested (one week between April 24 and April 30, 2012). Most of the result discrepancies were caused by additional retrieved documents provided by three contributors (Rijksmuseum; Koninklijke Bibliotheek; Open Beelden). In these cases, all additional documents were relevant to the queries but not found in the old ranking and the differences can be explained by these new additions.

Four queries (marked with a *) resulted in document differences even though the additional results were not provided from the three contributors mentioned above. It is important to point out that the additional documents found by the new ranking algorithm as compared to the old ranking were not relevant and may point towards errors in the processing of queries and documents. It is worth to look at them in closer detail.

Query and Document Processing Errors in the New Europeana Ranking Algorithm

Number stemming

The query “friday 13th” retrieved 138 more results in the new ranking, but the additional documents were non-relevant. Apparently, the new algorithm uses stemming that reduces the “13th" to a simple number “13”, which was not used in the old ranking (see Figure 35).

In order to generalize this finding, an analog query was created. Similar to the example of “friday 13th", the query “5th festival” also retrieved non-relevant documents based on the co-occurrence of “festival” and “5”, which seems to be caused by the stemming of “5th" to “5" (Figure 36]. In general, this seems to indicate an inappropriate number stemming.

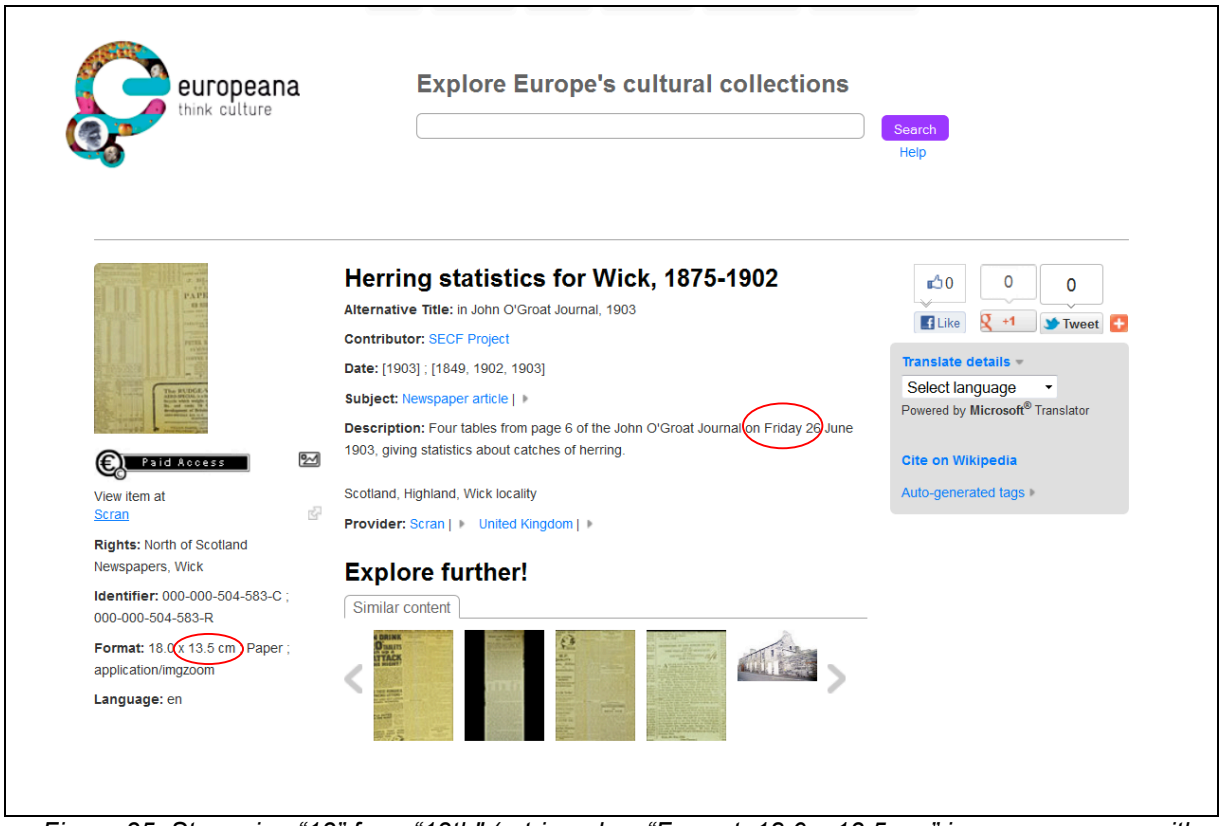

Figure 35: Stemming "13" from "13th" (retrieved as "Format: 18.0 x 13.5 cm" in co-occurrence with "Friday")

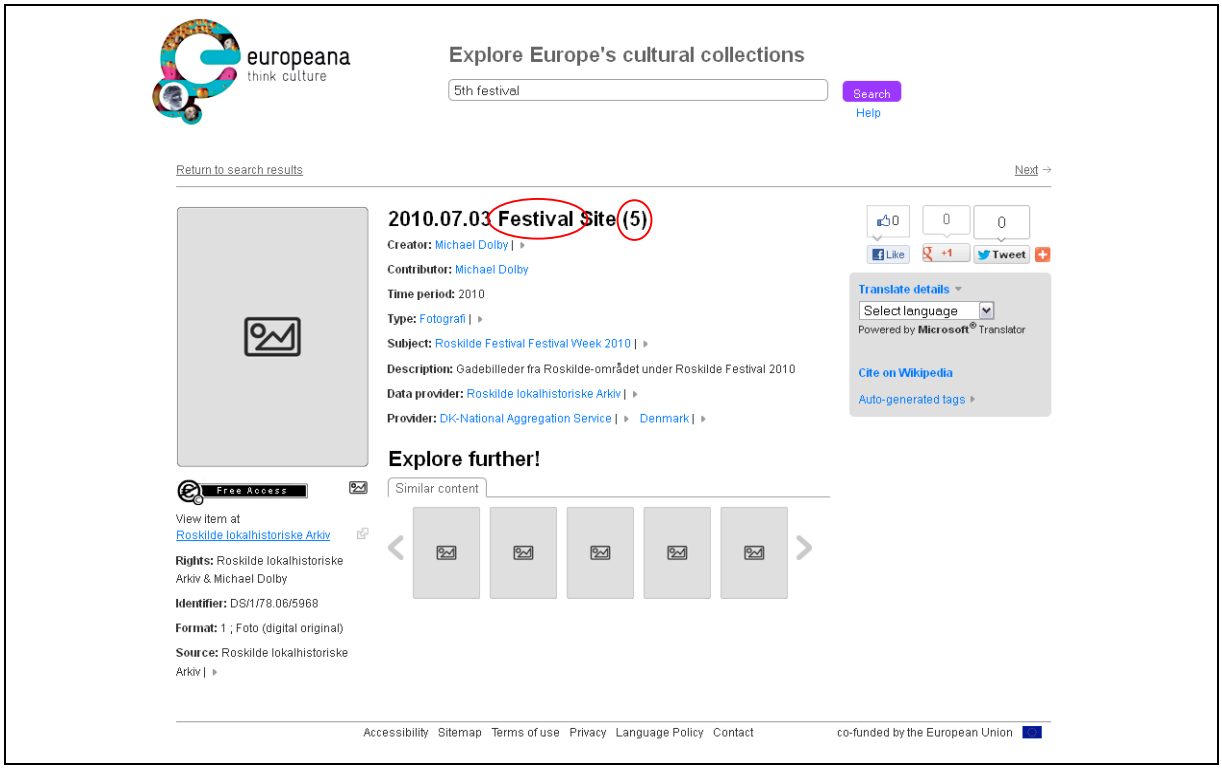

Figure 36: Inappropriate number stemming ("5th" to "5")

Hyphenation

Likewise, the query "jean-jaques rousseau" (misspelled in the original) retrieved 2 more documents in the new ranking. Regardless of the misspelled "jaques" instead of "jacques",

the reason for the additional retrieval results seem to be related to the hyphen (“-“) in the query.

In the old ranking, all retrieved documents showed the string “jean-jaques” or at least “jean jaques” whereas the new ranking also retrieved documents with the single strings “jean” and “jaques”, which could lead to results that had these names in different contexts and were not relevant to the original query. Figure 37 shows an example, where the de-hyphenation still results in a relevant document.

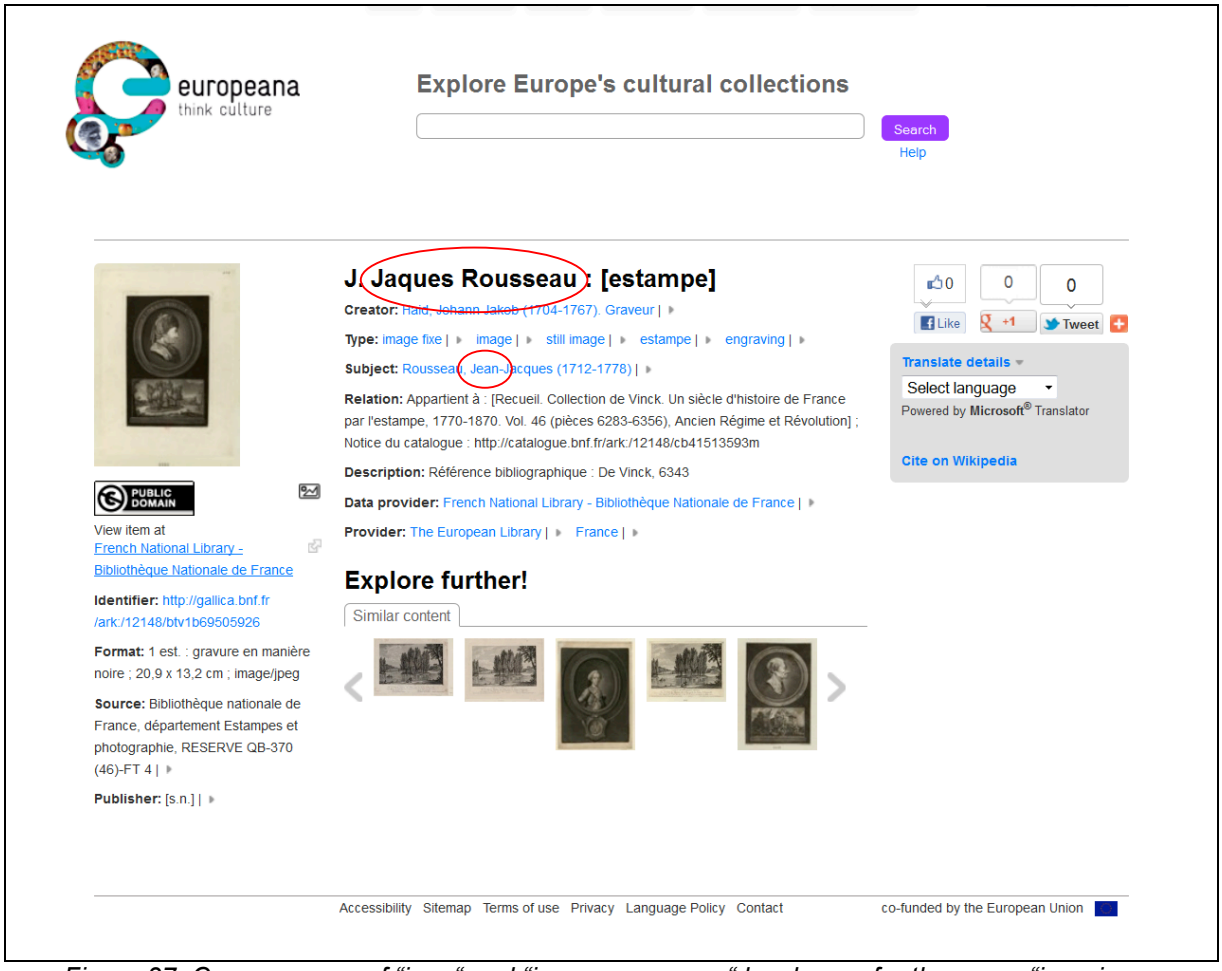

Figure 37: Co-occurence of “jean“ and “jaques rousseau“ by chance for the query “jean-jaques rousseau”

The other additional result in the new ranking showed a non-relevant co-occurrence of the strings “jean” and “jaques” and “rousseau” (not related to the name Jean-Jacques Rousseau) (Figure 38).

As an analogy the query “Gustav-Adolf Schreiber” retrieved also non-relevant results as co-occurrence of the single strings “Gustav” and “Adolf” and “Schreiber” (Figure 39). In general, this seems to indicate an inappropriate isolation of hyphenated compound names.

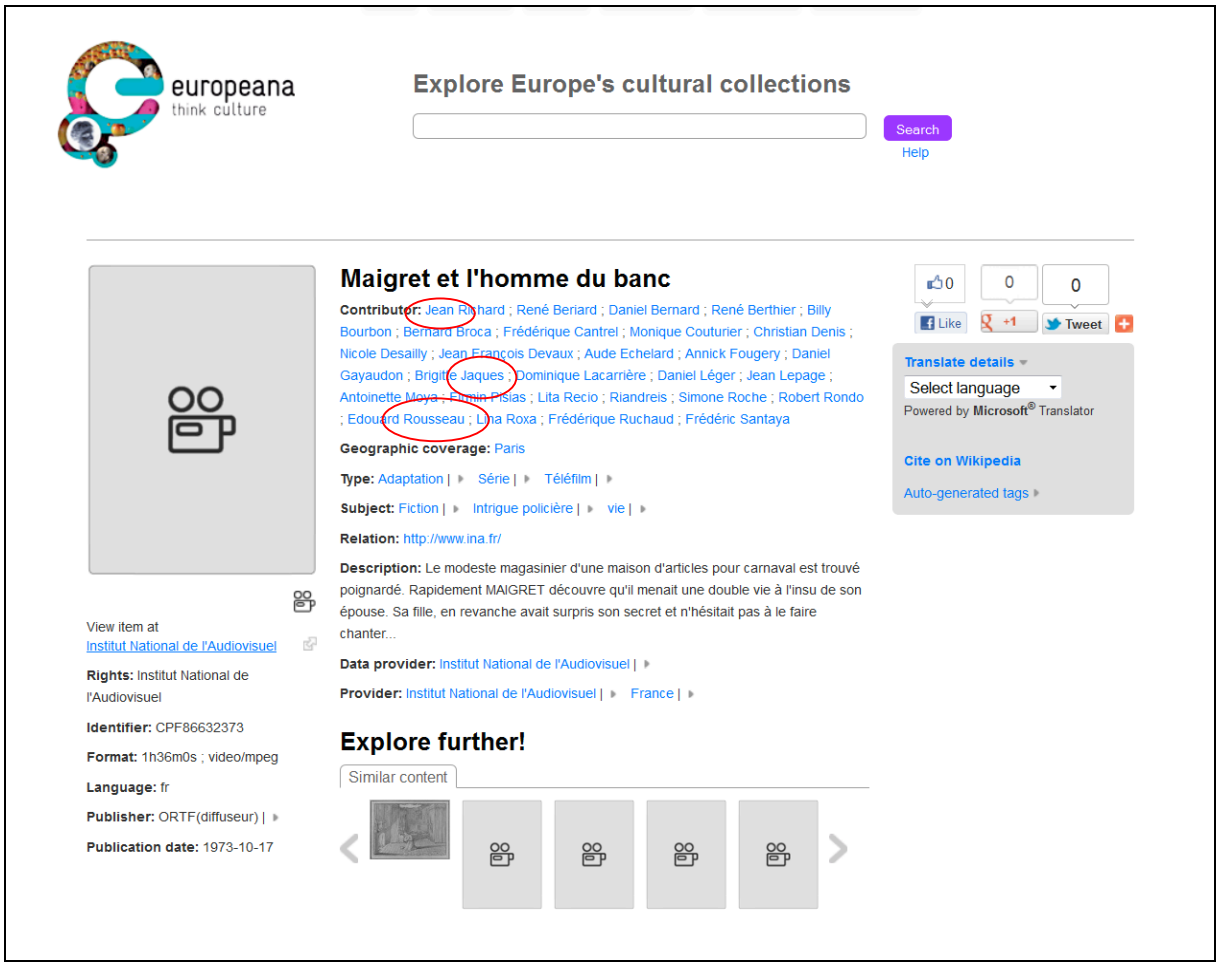

Figure 38: Non-relevant co-occurrence of "Jean" and "Jaques" and "Rousseau"

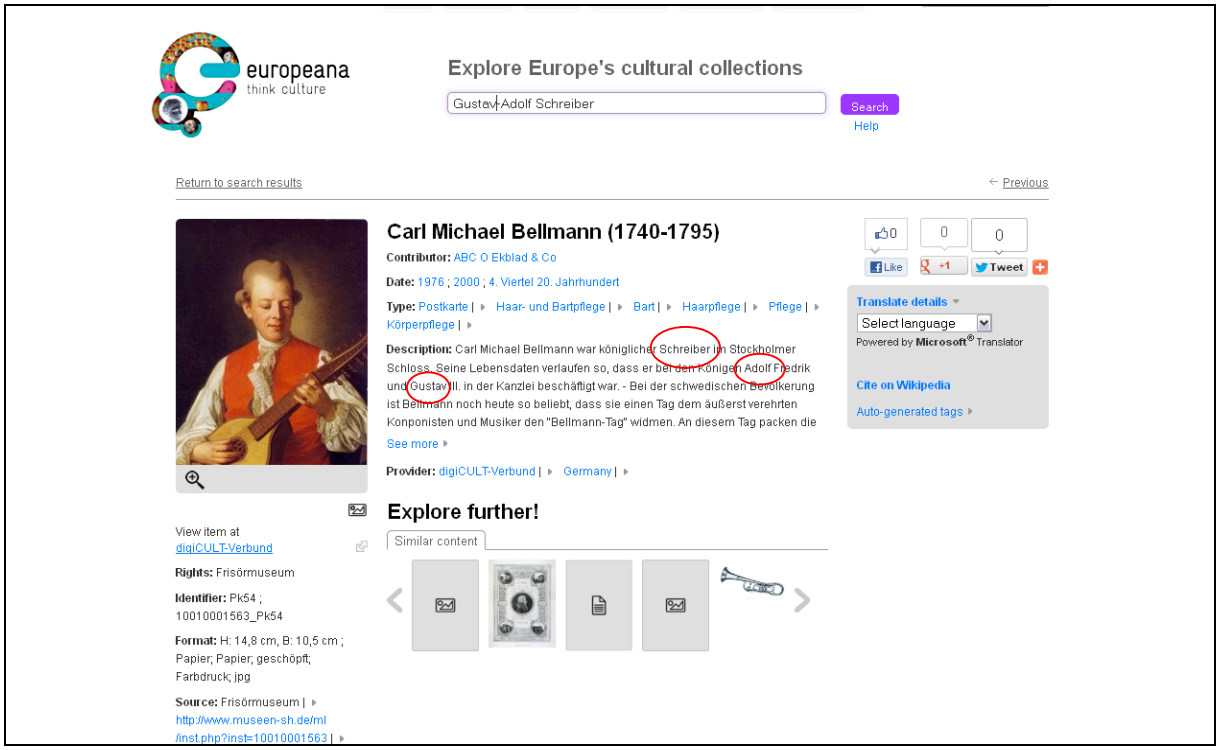

Figure 39: Inappropriate isolation of compound names with a hyphen ("Gustav-Adolf" to "Gustav" and "Adolph")

Text Stemming

The query “song medieval” retrieved one more document in the new ranking, which was also non-relevant. This seems to be related to a different stemming in the new ranking. The additional result showed a co-occurrence of “medieval” and “songe”, the latter obviously stemmed to “song” (Figure 40).

Figure 40: Co-occurence of “medieval” and “songe” for the query “song medieval”

As an analogy to the case “song medieval”, the query “muse” retrieved also non-relevant documents caused by a stemming of “musee” to “muse” – possibly a more grave error (Figure 41). Again, this seems to indicate an inappropriate stemming, which is cutting more than common suffixes.

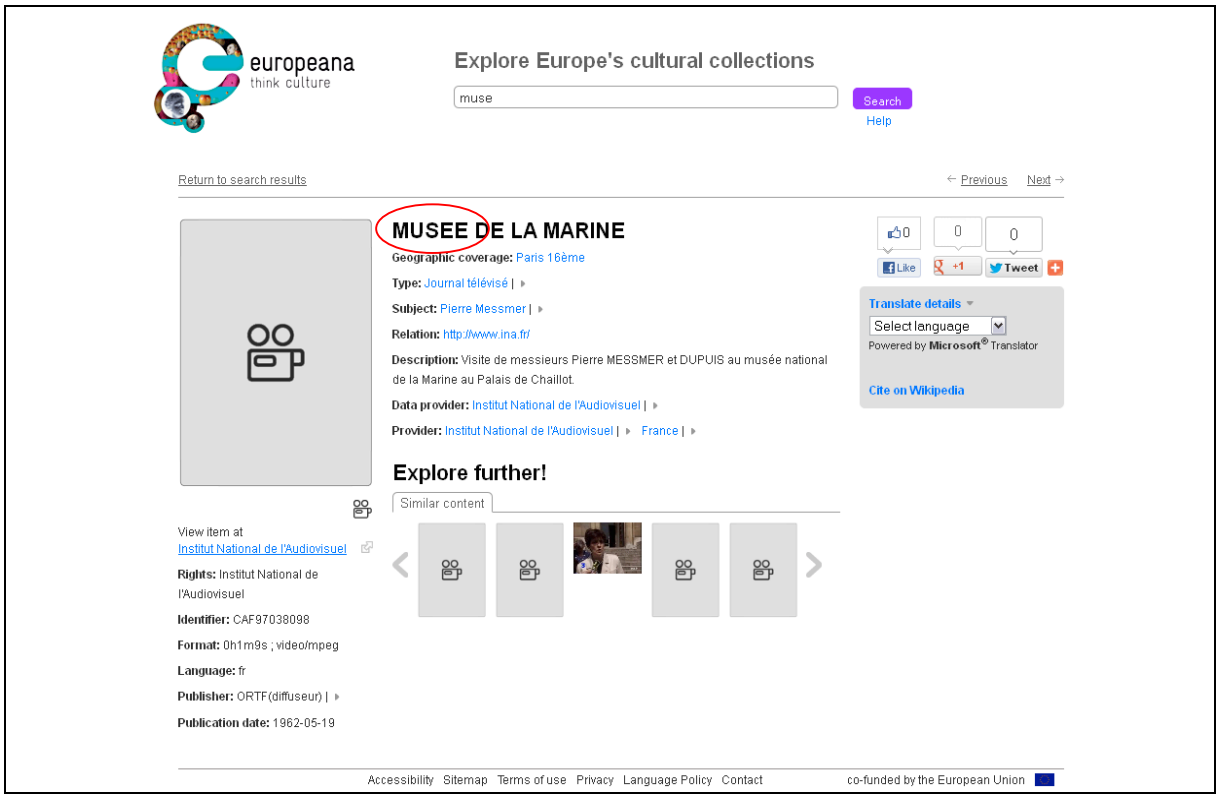

Figure 41: Inappropriate stemming ("musee" to "muse")

The Curios Case of "QUATREMERE"

The query "QUATREMERE" showed no overlap in result sets. It retrieved 23 different documents in the new ranking and 21 documents in the old ranking.

The retrieved documents in the old ranking showed the string "quatremere" or a compositum such as "Quatremere-Disjonval" in each case. In contrast, the retrieved documents in the new ranking showed only the two stemmed forms "quatremer" or "quarteme" (even the form "quatrem" retrieved the same 23 results).

The 21 relevant results of the old ranking are not retrieved in the new ranking but are still in the EUROPEANA collection. There seems to be no obvious reason for that discrepancy. There is no normalization of the misspelled form "quatremere" (no *accent grave*) to the standard form "quatremère" but this fact did not cause the difference. Even documents with the string "Quatremere" in the document title could not be retrieved with the query "QUATREMERE". In contrast, a combination such as "QUATREMERE de" or even "QUATREMERE d" retrieved these missing relevant documents (Figure 42). It is unclear how this error is caused and what other examples could be found.

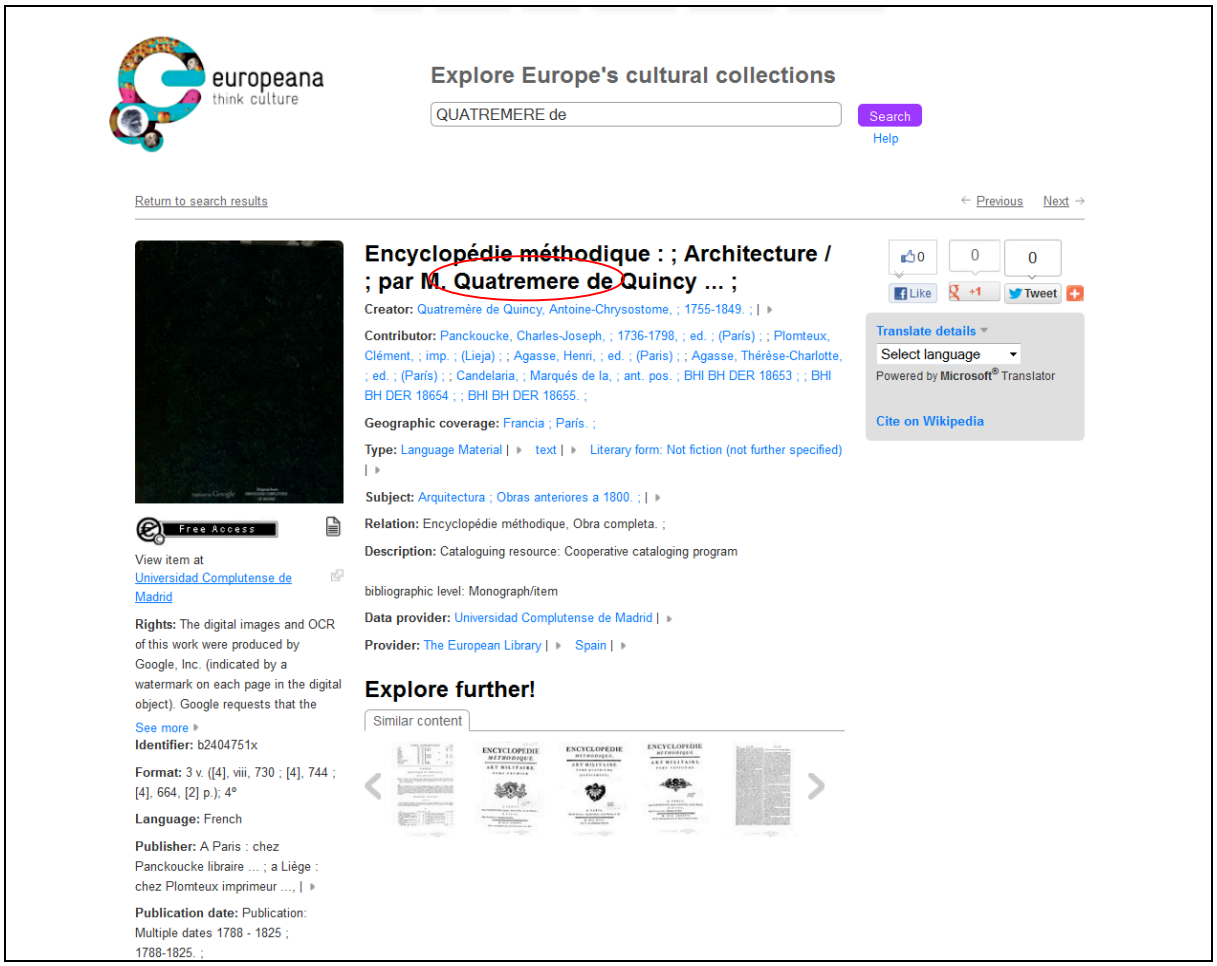

Figure 42: Relevant document not retrieved with query "QUATREMERE" in spite of the occurrence in the title

Result List Ordering / Different Ranking

Even if the same result sets (the same documents) are retrieved, different ranking algorithms might retrieve them in different order. The goal for the new ranking algorithm should be to improve the ranking, that is, the ordering of the results.

A first analysis showed that among the first 12 results per query (equals one Europeana result page), only 3.84 documents overlapped when the new and old ranking algorithm were compared. This does not yet consider, whether the first-ranked documents can be considered relevant for the user, which was the task of the CHiC experiments.

7.2 CHiC 2012 - Cultural Heritage in CLEF

The CHiC lab researched information retrieval systems for the cultural heritage environment by using real data, real user queries and real tasks. CHiC used the Europeana collections and Europeana queries (gained from log files) in order to compare different information retrieval approaches for this kind of cultural heritage data.

In order to address the specific requirements in cultural heritage environments, three evaluation tasks were created, an ad-hoc evaluation task, which was considered as a baseline for information retrieval systems, a semantic enrichment task and a variability task.

The ad-hoc task tested the standard information retrieval use case, where a single query is sent against the system and the system retrieves a set of relevant documents. The variability

task expanded the ad-hoc task by requiring the systems to retrieve diverse results within the first 12 results. Diverse results are not only relevant to the query, but should present a particular good overview over the different object types and categories targeted towards a casual user, who might like the "best" ones possibly sorted into "must sees" and "other possibilities." The semantic enrichment task asked for systems to suggest other terms and phrases to enrich the original query. This task tests whether a system is capable of supporting users in formulating appropriate search queries.

All three tasks were tested on 50 or 25 queries (taken from the original 85) and were offered in monolingual, bilingual (query language different from document language) and multilingual (documents in multiple languages) modes. For the bilingual and multilingual modes, the queries were translated into English, German and French to search documents in at least these three languages (and other documents that contained the same terms).

The documents retrieved from the information retrieval systems under test were judge for relevance according to the query they were retrieved for. The intellectually created relevance assessments were supported by the Distributed Information Retrieval Evaluation Campaign Tool (DIRECT)[22] system which is created for evaluation of multilingual information access in the project Participative Research labOratory for Multimedia and Multilingual Information Systems Evaluation (PROMISE)[23].

More information on the lab in general and the results and outcomes can be found in the overview paper (Petras et al., 2012) and the individual working note papers[24].

Participants of CHiC 2012 evaluation lab reported that the specific challenges in the cultural heritage domain using EUROPEANA data were qualitatively diverse metadata, a high amount of named entities search, very short queries and vague information needs.

Europeana Rankings at CHiC ad-hoc Task

Within the context of the multilingual ad-hoc task, the result sets for the 50 queries from the old and new Europeana ranking algorithm were submitted for judgement. Result metrics are always averaged over all 50 queries. On average, a query would result in 109 relevant documents.

Mean average precision is an information retrieval metric that is usually used to compare between system performances because it has been shown that it achieves robust results. Precision measures how many documents of the result set are relevant. This metric also considers the position of the relevant documents. Mean average precision takes the precision score of each relevant document in a result list for the query and averages it over all documents (average precision). For 50 queries, the average precision is then again averaged (mean average precision).

Whereas the old Europeana ranking achieves a mean average precision of 20.05%, the new Europeana ranking achieves a mean average precision of 23.02%. The new ranking shows a 15% improvement, which is commonly considered significant. Because of missing comparisons, we didn’t perform a significance test.

Precision at 10 documents is another popular information retrieval metric. It measures how many relevant documents are in the first 10 documents, usually a metric to compare how

[22] http://direct.dei.unipd.it/

[23] http://www.promise-noe.eu/

[24] http://clef2012.org/index.php?page=Pages/proceedings.php

good information retrieval systems are on the first page. The old Europeana ranking achieves a precision at 10 of 36.3%, which means roughly one third of the first shown results are relevant. The new Europeana ranking achieves a precision at 10 of 37.4% documents. For this measure, the two ranking algorithms show no significant differences between each other, meaning that both ranking algorithms retrieve about the same number of relevant documents in the first 10 results.

These two short comparisons show that the two ranking algorithms might not be as different as hoped. However, the new ranking algorithms showed an improvement in one metric (mean average precision), demonstrating that it probably will retrieve more relevant results in general.

The CHiC 2013 evaluation lab will again perform a multilingual ad-hoc task, where several systems will compete in achieving the best performance for multilingual retrieval. The new lab will not only take three languages into account but all languages that Europeana documents are offered in. The results achieved in the new lab can be again compared with the Europeana ranking performance and more points for improvement found.

8. Conclusion

The preparation and execution of the CHiC 2012 lab using EUROPEANA data have been shown that the cultural heritage domain requires specific evaluation activities. In particular, the high amount of named entity search suggests the use of a named entity recognizer for a better handling of stemming and to focus future work on topical queries. The challenge of very short queries and vague information needs require more knowledge about the user's context, information seeking behavior and interaction patterns. In general, there is a need for more EUROPEANA training data. The differences between the old and new EUROPEANA ranking algorithm have been shown an improvement of recall, but indicate further challenges in precision. In comparison to the old ranking many non-relevant documents were retrieved caused by an inappropriate stemming such as the isolation of compound names with a hyphen, or a too strong number and suffix stemming.

9. Results and Future Work

This deliverable described the different aspects of multilingual access to digital cultural heritage content. It analysed different cultural heritage websites with regard to their multilingual access strategy. The analysis showed that many of them already offer multilingual content and provide interfaces in different languages. Currently, none of them supports search and browsing functionalities across languages. Nevertheless, there is a trend to offer browsing and discovery tools, which enable users to explore content without the need to formulate a query. Some of these features such as map and timeline browsing allow the user to retrieve documents in languages she might not understand.

With regard to the language diversity of the content and the multilingual audience, Europeana is unique and can act as a trailblazer in providing multilingual access and novel interaction models to explore multilingual content. It was shown that Europeana already offers many multilingual access points. Major achievements are the multilingual enrichments of the metadata that facilitate retrieval across languages and the curated exhibitions, which highlight content in several languages. To improve these features and offer seamless multilingual access, some challenges need to be faced. Some are easy wins that can overcome confusion on the user side by providing more help texts. Others intervene with the search workflow introducing more clicks and cognitive efforts on the user side.

As shown in chapter 6, there are possibilities to leverage user-driven translations provided during the search process. Features like this quickly get very complex and run the risk of clattering the user interface detracting the user form his tasks. The same is true for collaborative features such as social tagging. They require complex user management systems and strategies for maintaining and displaying tags in different languages. Strategies for such features also need to include considerations about the incentives offered for users to participate and the transparent purpose the tags will serve.

Furthermore, users turn to Europeana more and more on mobile devices. Strategies for multilingual interactions and display of content in diverse languages also need to include the limitations of smaller screen and different interaction habits such as swiping the screen.

In the second half of the project, use cases and interaction models will be provided. They will be specifically targeted on enriching objects with multilingual user-driven data and how to present that to the user. In a second step, a staged model will be developed which presents possibilities of enriching the object's metadata multilingually during the ingestion process.

10. References

AGOSTI, M., CRIVELLARI, F., DEAMBROSIS, G., FERRO, N., GÄDE, M., PETRAS, V. & STILLER, J. 2009. D2.1.1 Report on User Preferences and Information Retrieval Scenarios for Multilingual Access in Europeana. EuropeanaConnect.

BILAL, D. & BACHIR, I. 2007. Children's interaction with cross-cultural and multilingual digital libraries. II. Information seeking, success, and affective experience. *Information Processing & Management,* 43 65-80.

BOER, V. D., ISAAC, A., SCHREIBER, G., OSSENBRUGGEN, J. V., WIELEMAKER, J. & STILLER, J. 2011. D2.3.1 Multilingual mapping of schemes and vocabularies. EuropeanaConnect.

BOSCA, A. & DINI, L. The Role of Logs in Improving Cross Language Access in Digital Libraries. Proceedings of the International Conference on Semantic Web and Digital Libraries, 2009.

CALLAHAN, E. S. & HERRING, S. C. 2011. Cultural bias in Wikipedia content on famous persons. *Journal of the American Society for Information Science & Technology,* 62**,** 1899-1915.

CLARK, D. J., NICHOLAS, D. & ROWLANDS, I. 2011. D3.1.3 – Report on best-practice and how users are using the Europeana service. Europeana: an evaluation of users, usage, and information-seeking behaviour derived from the webserver log-files of europeana.eu (October 2009–October2011). EuropeanaConnect.

CLEMMENSEN, T. & ROESE, K. 2010. An overview of a decade of journal publications about Culture and Human-Computer Interaction (HCI). *Human Work Interaction Design: Usability in Social, Cultural and Organizational Contexts***,** 98-112.

COOPER, A., REIMANN, R. & CRONIN, D. 2007. *About face 3: the essentials of interaction design*, Wiley Pub.

CRUMLISH, C. & MALONE, E. 2009. *Designing Social Interfaces,* Sevbastopol, O'Reilly.

DEKKERS, M., GRADMANN, S. & MEGHINI, C. 2009. Europeana Outline Functional Specification: D 2.5 – Europeana Thematic Network Project.

DIX, A. 2004. *Human-Computer Interaction*, Prentice Hall.

DOBREVA, M. & CHOWDHURY, S. 2010. A User-Centric Evaluation of the Europeana Digital Library. *In:* CHOWDHURY, G., KOO, C. & HUNTER, J. (eds.) *The Role of Digital Libraries in a Time of Global Change. ICADL 2010.* Springer Berlin / Heidelberg.

ELETA, I. & GOLBECK, J. A study of multilingual social tagging of art images: Cultural bridges and diversity. 2012 Seattle, WA. 695-704.

FERRO, N. & PETRAS, V. 2009. MLIA4DL - Multilinguality in Information Access to Digital Libraries: User Needs and Evaluation of multilingual resources use. *Workshop at the International Conference on Digital Libraries and the Semantic Web 2009 (ICSD2009).* Trento, Italy.

FORD, G. & GELDERBLOM, H. The effects of culture on performance achieved through the use of human computer interaction. 2003. South African Institute for Computer Scientists and Information Technologists, 218-230.

GAO, W., NIU, C. & NIE, J.-Y. Cross-lingual query suggestion using query logs of different languages. 30th annual international ACM SIGIR conference on

Research and development in information retrieval, 2007 New York, NY, USA. ACM, 463−470.
GHORAB, M. R., LEVELING, J., ZHOU, D., JONES, G. J. F. & WADE, V. 2010. TCD-DCU at logCLEF 2009: An analysis of queries, actions, and interface languages. *10th Workshop of the Cross-Language Evaluation Forum, CLEF 2009.* Corfu, Greece.
GONZALO, J., PEÑAS, A., VERDEJO, F. & PETERS, C. 2008. Workshop on Best Practices for the Development of Multilingual Information Access Systems: the User Perspective: D3.2 - TrebleCLEF-Project.
HE, D., WANG, J., OARD, D. & NOSSAL, M. 2003. Comparing user-assisted and automatic query translation. *Advances in cross-language information retrieval*, 400-415.
HEARST, M. 2009. *Search user interfaces,* Cambridge ; New York, Cambridge University Press.
HOFSTEDE, G. H. & HOFSTEDE, G. 2001. *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations*, Sage Publications, Inc.
HOWE, J. 2010. Crowdsourcing: a defintion. *Crowdsourcing* [Online]. Available from: http://crowdsourcing.typepad.com.
IRN-RESEARCH 2009. Europeana online visitor survey - Research Report Version 3. IRN Research.
ISAAC, A. 2010. *Functional Requirements: Data Enrichment.* [Online]. Available: http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsEDMDataEnrichment.
ISAAC, A. 2011. *EDM Prototyping: 2.1. Enrichment of EDM data.* [Online]. Available: http://www.europeanalabs.eu/wiki/EDMPrototypingTask21.
KRALISCH, A., EISEND, M. & BERENDT, B. 2005. The impact of culture on website navigation behaviour. *Proc. of the 11th International Conference on Human-Computer Interaction.*
LARGE, A. & MOUKDAD, H. 2000. Multilingual access to web resources: an overview. *Program: electronic library and information systems,* 34, 43-58.
LOPEZ-OSTENERO, F., GONZALO, J. & VERDEJO, F. 2005. Noun phrases as building blocks for cross-language Search Assistance. *Information Processing & Management,* 41, 549-568.
OARD, D., GONZALO, J., SANDERSON, M., LÓPEZ-OSTENERO, F. & WANG, J. 2004. Interactive cross-language document selections. *Information Retrieval,* 7, 205-228.
OARD, D. W. A Comparative Study of Query and Document Translation for Cross-Language Information Retrieval. Third Conference of the Association for Machine Translation in the Americas on Machine Translation and the Information Soup, 1998.
OARD, D. W. & DIEKEMA, A. R. 1998. Cross-Language Information Retrieval. *Annual Review of Information Science and Technology,* 33, 223-256.
OARD, D. W., HE, D. & WANG, J. 2008. User-assisted query translation for interactive cross-language information retrieval. *Information Processing & Management,* 44, 181-211.
OLENSKY, M., STILLER, J. & DRÖGE, E. Poisonous India or the improtance of a semantic and mutlilingual enrichment strategy. 2012 Cádiz, Spain.
PETERS, C., BRASCHLER, M. & CLOUGH, P. 2012. *Multilingual Information Retrieval: From Research to Practice,* Heidelberg, Germany, Springer.

PETRAS, V. 2011. D2.7.1 - Report on Multilingual Access Strategies to Digital Libraries.

PETRAS, V., FERRO, N., GÄDE, M., ISAAC, A., KLEINEBERG, M., MASIERO, I., NICCHIO, M. & STILLER, J. 2012. Cultural Heritage in CLEF CHiC Overview 2012 (CLEF Online Working Notes/Labs/Workshops). *CLEF 2012.*

PETRELLI, D., DEMETRIOU, G., HERRING, P., BEAULIEU, M. & SANDERSON, M. 2003. Exploring the effect of query translation when searching cross-language. *Lecture Notes in Computer Science.* Berlin: Springer.

RANIERI, M., EMANUELE, P. & BENTIVOGLI, L. Browsing Multilingual Information with the MultiSemCor Web Interface. LREC 2004 Satellite Workshop on "The Amazing Utility of Parallel and Comparable Corpora", 2004. 4.

SMITH-YOSHIMURA, K. & SHEIN, C. 2011. Social Metadata for Libraries, Archives and Museums Part 1: Site Reviews. Dublin, Ohio: OCLC Research.

WU, D., HE, D. & XU, X. 2012. A study of relevance feedback techniques in interactive multilingual information access. *Library Hi Tech,* 30**,** 523-544.

11. Appendix

Archives Portal Europe	http://www.archivesportaleurope.eu/Portal/index.action	Aggregator	Metadata records	ES	Provider language
Beeld en Geluid	http://www.beeldengeluid.nl/	Archive	Video, Sound	NL	NL
Brooklyn Musem	http://www.brooklynmuseum.org/	Museum	Image	US	Interface language
DaheshMseum of Art	http://www.daheshmuseum.org/	Museum	Image	US	EN
David Rumsey Map Collection	http://www.davidrumsey.com/	Collection	Image	US	EN
Gallica	http://gallica.bnf.fr/?lang=EN	Aggregator	Text, Image	FR	FR
Google Art project	http://www.googleartproject.com/	Aggregator	Image	US	Interface language
HathiTrust	http://www.hathitrust.org/	Aggregator	Metadata records	US	EN
Historypin	http://www.historypin.com	Community	Image	US	Provider Language
ICDL - International Children's Digital Library	http://en.childrenslibrary.org/	Collection	Text	US	Multiple Languages
LibraryThing	http://www.librarything.de/	Community	Metadata records	US	Provider Language
Louvre	http://www.louvre.fr/	Museum	Image	FR	EN, FR

MaritiemDigitaal	http://www.maritiemdigitaal.nl/	Museum	Image, Text	NL	Provider Language
NationaalArchief	http://www.nationaalarchief.nl/	Archive	Image, Metadata records	NL	NL
NationaalHistorisch Museum	http://www.innl.nl/	Museum	Image, Text	NL	NL
Open Images	http://www.openimages.eu/	Archive	Video, sound	NL	EN, NL
Perseus Digital Library	http://www.perseus.tufts.edu/hopper/	Library	Text	US	EN
Philaplace	http://www.philaplace.org/	Collection	Video, Image	US	EN
Polar bear exhibtion	http://quod.lib.umich.edu/p/polaread/	Collection	Image, Text	US	EN
Project Gutenberg	http://www.gutenberg.org/wiki/Main_Page	Library	Text	US	Provider Language
Rijksmuseum	http://www.rijksmuseum.nl/	Museum	Image, Metadata records	NL	NL
ShelfLife DPLA Demo	http://librarylab.law.harvard.edu/dpla/demo/app/	Library	Text, Metadata records	US	EN
StädelMusem	http://www.staedelmuseum.de/	Museum	Image	DE	DE
Steve Tagger	http://tagger.steve.museum/	Museum	Image	US	EN
The Athenaeum	http://www.the-athenaeum.org/	Community	Image	UK	EN

The British Museum	http://www.britishmuseum.org/	Museum	Image	UK	EN
The European Library	http://www.theeuropeanlibrary.org/tel4/	Aggregator	Text, Metadate records	NL	Provider Language
The Frick Collection	http://collections.frick.org/	Museum	Image	US	EN
The State Hermitage Museum	http://www.hermitagemuseum.org/	Museum	Image	RU	Interface language
Victoria and Albert Museum	http://collections.vam.ac.uk/	Museum	Image	UK	EN
World Digital Library	http://www.wdl.org/en/	Aggregator	Text, Image	US	Interface language
Your paintings	http://www.bbc.co.uk/arts/yourpaintings/	Aggregator	Image	UK	EN