



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/4518/>

Conference or Workshop Item:

Clough, P., Marlow, J. and Sanderson, M. (2006) Designing multilingual information access to Tate Online. In: Workshop on New Directions In Multilingual Information Access, run at the ACM SIGIR Conference, 2006, August 06 -11, 2006, Seattle, USA.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

promoting access to White Rose research papers



Universities of Leeds, Sheffield and York
<http://eprints.whiterose.ac.uk/>

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/4518/>

Conference paper

Clough, P., Marlow, J. and Sanderson, M. (2006) *Designing multilingual information access to Tate Online*. In: Workshop on New Directions In Multilingual Information Access, run at the ACM SIGIR Conference, 2006, August 06 -11, 2006, Seattle, USA.

Designing Multilingual Information Access to Tate Online

Paul Clough
University of Sheffield
Western Bank
Sheffield, UK
+44 114 222 2664

p.d.clough@sheffield.ac.uk

Jennifer Marlow
University of Sheffield
Western Bank
Sheffield, UK
+44 114 222 2664

ldp05jm@sheffield.ac.uk

Mark Sanderson
University of Sheffield
Western Bank
Sheffield, UK
+44 114 222 2648

m.sanderson@sheffield.ac.uk

ABSTRACT

The Tate is Britain's premier national art gallery and includes content from internationally-renowned artists such as Constable and Turner. Like most cultural heritage institutions, the Tate provides online access to a large amount of digitized material. Given the international importance of content provided by the Tate Gallery, multilingual access would seem an ideal way in which to increase accessibility to the collections, and thereby increase traffic to the website. In this short paper we propose using the Tate as a case study for cross-language research and evaluation, determining the gallery's requirements and the multilingual needs of their end-users.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]

Keywords

Cross-Language IR, localization, internationalization

1. INTRODUCTION

The Tate is Britain's national gallery and houses both the national collection of British art from the 16th Century and international modern art. The Tate has one collection shared between four physical galleries: Tate Britain, Tate Modern, Tate Liverpool, and Tate St Ives. Each gallery has an online presence drawing from the same database or collection of digitized content (over 65,000 works in the collection). The Tate Modern Gallery is the world's most visited museum of contemporary art, with over 4 million visitors per year. Similarly, the Tate Online¹ website attracted over 7 million unique visitors in 2005. It is the most popular UK visual arts and museum website.

Currently, most of the content on the Tate's website is only available in English. This is somewhat ironic, given the museums' international reputation and reach. Although a lack of resources makes it impossible to create fully localized versions of the site in other languages, it is safe to say that a wider audience could be reached if more multilingual content was available or users could perform cross-language searches. Most of the current multilingual content on Tate Online is geared towards individuals who are making a physical visit to one of the galleries (i.e. it provides information on directions, opening and closing times, etc.). However, the Tate website is now becoming a destination in its own right and therefore of interest to a global audience².

¹ <http://www.tate.org.uk/>

² The Collections section of the Tate Online is of particular international interest.

In this short paper, we describe Tate Online and discuss work in progress for assessing their multilingual requirements and the needs of end-users. This will act as the basis for future work with the Tate, the ultimate goal being the *implementation* of a multilingual version of Tate Online. Given the current interest in access to cultural heritage resources (particularly in Europe), we believe that cross-language research and evaluation can learn from "real-world" examples like Tate Online. The commercial constraints and the challenges of producing such applications could be used as parameters in which to run future multilingual evaluation (e.g. based on time, cost, manpower effort etc.). By considering not only the technical aspects of cross-language and the requirements of end-users, but also business needs may help to reduce the gap between cross-language research and its use in practice.



Figure 1. Example caption for a Collection object (copyright of the Tate).

2. BACKGROUND

2.1 Tate Online

Tate Online provides a wide variety of information about the physical galleries including events, directions to the gallery, personnel and on-site shops and services. The gallery sites attract large numbers of visitors annually from around the world, serving as popular tourist attractions. The Tate holds a large collection of works, the galleries publicly displaying some of the more popular exhibits. Much of the content from the Tate is available online; especially the collection and more people visit Tate Online than the physical gallery itself. Entry to the Tate is free but income is generated from charging for access to certain exhibitions,

membership fees, and shops and restaurants. One of the purposes of Tate Online is to generate revenue by attracting visitors to the Tate galleries.

2.2 The Tate collection

The Tate's collection contains over 65,000 works, held by the Tate and made available to users online. Works in the collection contain a variety of descriptive metadata varying from a few words to biographical descriptions and display captions for more popular works (Figure 1). The website supports full-text searches, simple searches (restricted to the artist's name or the name of the work) and advanced searches (includes filters for types of objects and dates). In addition to search, the collection can be browsed via an Artist A-Z index (e.g. A Abbey to Ayrton; B Bacon to Buvelot), a hierarchical subject search (e.g. Architecture > monuments > cross) and an illustrated glossary (e.g. Abstract art; Abstract expressionism; Actionism).

2.3 The Tate and Multilingualism

Rellie outlines two primary goals for UK museums online in 2006 [1]: increasing audiences and improving access. One important way of achieving this is to make website content available in many languages. Tate Online recognizes this need and has set a priority of broadening, expanding and diversifying its real and virtual audiences [2] while pursuing the translation of content as part of a greater mission to localize digital learning [3]. This would benefit both the website visitor and the Tate, as "the localization of products and services based on an awareness of and response to linguistic, cultural and technical differences is seen to bring a substantial return on investment" [3].

Table 1. Geographical distribution of Tate Online users.

Region	%
UK	68.8
USA	3.9
Belguim	3.6
France	3.5
Sweden	2.0
Netherlands	2.9
Denmark	1.8
Italy	1.1
Spain	1.0
Canada	1.1
Australia	1.5
Other	8.8

In 2005 more people visited Tate Online than all four of the physical galleries combined [1]. Most traffic to the site is split fairly evenly across three areas: pages with information about the physical gallery, the online collection, and the shop/e-learning sections [1]. Of these areas, the collections section is of greatest interest for translation because it represents the core of the museum. This resource is a fully digitized collection of the Tate's 65,000 works of art, and currently it is only searchable in English. Table 1 shows the geographical distribution of users from a 2004 online survey by the Tate, indicating that 24% of users are from outside the UK, Canada, USA and Australia. One aim is to increase this number through localization.

The Tate has no multilingual resource available in-house. Currently, a limited amount of information about the Tate and

visiting the physical galleries has been manually translated into a number of languages including German, French, Arabic and British Sign Language (BSL). This work was funded by an external funding agency and outsourced to a commercial translation company (although Greek and Polish versions were translated voluntarily).

2.4 Existing websites

Two existing online museum websites also providing important artistic content are the Rijksmuseum³ and the Van Gogh Museum⁴. The Rijksmuseum provides separate English and Dutch versions of the website. Many of the works of art have bilingual captions enabling users to search the collection in Dutch or English, without requiring query translation. The Van Gogh museum is more multilingual and enables the user to select between English, Dutch, Spanish, French, Italian and Japanese. The majority of content is accessible in English and Dutch, with a selection of information translated into the other languages and search does not appear to use query translation.

3. RELATED WORK

3.1 Designing localised websites

Adapting websites to meet the linguistic and cultural needs of the local communities they target is referred to as *globalisation*. The different versions are known as *localised* websites and often require specific design considerations [4-7]. These might include: which languages to translate and how many, an awareness of cultural issues (e.g. the use of specific terminology or offensive references), the availability of resources (e.g. manpower, translation tools), technical and maintenance issues, how to measure success and issues surrounding design. The W3C [4] differentiate between *international* and *multilingual* websites: the former a website which is intended for an international audience; the latter a website which uses more than one language. In this definition, a multilingual site is also concerned with regional and cultural differences in addition to language, and international sites are often multilingual, e.g. a global company with information written in different languages.

The Eurescom guidelines for building multilingual websites [5] discuss the division between locale-dependent and locale-independent information common to most websites including: formatting of characters and layout, content (text, graphics etc) and navigational information (e.g. menus). Multilingual versions of a website may also exhibit different degrees of parallelism, ranging from a collection of monolingual sites at one extreme to a completely parallel site with identical structure, navigation and content at the other. Typically a trade-off must be made between the cost and effort involved in creating such a site and its benefit. Further issues to consider include [5]: the use of static versus dynamic content and whether off-line processing can be used to generate multilingual content, the use of Cross-Language Information Retrieval (or CLIR) for query translation, the translation of content and the role/use of automatic versus manual translation.

3.2 CLIR

CLIR deals with retrieval of documents written in one language by a query written in another [8]. Retrieval is achieved by

³ <http://www.rijksmuseum.nl>

⁴ <http://www3.vangoghmuseum.nl/>

translating queries or documents (or both) into the same (or a common) language and then applying standard monolingual retrieval. Classically, translation methods were divided into three approaches: (1) using bilingual dictionaries, (2) extracting word/phrase equivalents from parallel or comparable corpora, and (3) using a Machine Translation (MT) system. Although it is important to note that these three don't represent three different approaches: MT systems use a super-set of approaches using bilingual dictionaries, making extensive use of statistical models derived from parallel corpora as well as using parsing..

Gonzalo and Peters [9] discuss the effectiveness of current multilingual search systems from results of the Cross-Language Evaluation Campaign (CLEF). They show that between 1997 and 2002, results for the best English-French systems have increased MAP from 49% of monolingual effectiveness to 83%, and for English-German systems from 64% to 86%. Even for less studied language pairs, such as Italian-Spanish and French-Dutch, the effectiveness is at least 82% of monolingual performance.

MT systems enable multilingual access by providing translation of not only search requests, but also translation of documents (e.g. reports, Web pages etc.). Although many have been sceptical of MT systems in the past, Hutchins [5] comments that "the aim is now to produce aids and tools for professional and non-professional translation which exploit the potentials of computers to support human skills and intelligence, or which provide rough translations for users to extract the essential information from texts in foreign languages." MT is attractive due to the ease with which translation can be performed.

In addition to methods of translation, considering how real users interact with cross-language systems is also important and will influence their design (see, e.g. [11]). Studying the user and their search behavior results in developing systems which are more likely to succeed and be used in practice.

4. PROPOSED RESEARCH

4.1 Overview

The goal of this work is to establish the feasibility and benefits of providing users with multilingual access to the Tate's online content. A set of recommendations concerning a strategy for making the current Tate Online sites more useful and beneficial to individuals who do not speak fluent English will be generated and tested. This will involve gathering information about the most important needs individuals have when navigating the Tate's website, so that the best way of improving the value of current content may be ascertained. We will work with members of the Tate and end-users of Tate Online to establish requirements and assess the feasibility of multilingual access using qualitative and quantitative research methods. The work is structured into the following tasks:

- Review existing museum websites that exhibit different language versions to accessible content.
- Understand the business context of the Tate, e.g. their competition, position in the field and available resources, e.g. time, money, manpower.
- Identify current users of Tate Online and establish their multilingual needs and/or preferences, e.g. for search, navigation and presentation of static/dynamic content.
- Investigate the availability of multilingual resources and estimate their effectiveness on sample data.

- Given the needs of the end-users and the Tate, ascertain which parts of the Tate should be locale-dependent and locale-independent (i.e. what is translated and how).

Resources offered by the Tate include: access to log files, interviews with members of staff and the distribution of surveys to current users. The Tate have search logs which can be used to compile statistics about visitors to the Tate Online, such as where in the world they are coming from, and which pages they visit. Based on information extracted from the log files, a questionnaire will be created which will survey international visitors about the most important aspects of Tate Online to have available in a language other than English. This will help to determine which areas translation resources should be focused and prioritized. The questionnaire will be made available online, accessible through a link or banner placed on the Tate's website.

Results of this survey will form the basis for creating possible mock-up site interfaces based on a range of translation resources. A usability study is planned using actual museum visitors as participants. The Tate has run similar usability studies in the past and therefore has the equipment and the resources to make this a possibility. We also plan to translate a number of realistic queries extracted from the log file to test various approaches of query translation.

Throughout the process, one aspect to consider is whether international visitors viewing the Tate site actually desire or require more translated content. Using machine-translation software to do the physical translation work has the advantage of being more efficient and using fewer resources. However, it is not perfect and may produce errors, thus detracting from the overall professional and credible appearance of the site. Therefore, it is important to identify those areas that are in most need of translation and test them to ensure quality comes before quantity. Following recommendations in [7] and users of Tate Online, we will determine the locale-dependent and locale-independent parts of the website, including formatting, content of static and dynamically-generated pages, navigational aids such as menus, glossaries and other terminological lists, and query translation.

Finally, we will advise the Tate on what areas of the site are most important to have translated, what languages are most necessary, and the means by which the translation should be done. The end goal of the research will be to give the Tate a plan for how to direct limited resources, both now and in the future, to improve the value of their site for visitors who speak little or no English (if such users exist).

4.2 Constraints and limitations

The degree of multilingual access provided will ultimately depend on resources available to the Tate. This is true for any commercial application which research often overlooks. For example, the Tate has specified the following amounts available for translation and development of their website based on different sources of funding: (1) Entry-level: funded in-house up to £5,000; (2) Lottery-funded: funding up to £50,000; (3) European lottery-funded: up to £500,000. Of course, the funding required depends on the outcome of this feasibility study, in particular the trade-off between cost and benefit to the Tate and its end-users.

5. DISCUSSION

It is widely acknowledged in IR that large-scale evaluation campaigns such as TREC and CLEF offer valuable data for researchers to test new algorithms and methods. Such evaluations are also useful to companies as they provide a resource for current IR research and an estimate of how successful different methods are likely to be.

However, inspired by our experiences so far with Tate Online, it is clear that cross-language search (typically the focus of evaluation thus far) is only one aspect to providing multilingual access. In addition to tasks which evaluate multilingual search, we argue that tasks which also assess the effectiveness of other aspects of multilingual access such as translating the interface and results pages would help to identify limitations with existing translation resources and provide quantitative data for companies such as the Tate to help make decisions on regarding approaches for implementing multilingual access. In addition, evaluation must take into account the end-user, but also consider the practical constraints of a real-world environment where resources such as time, money and manpower are also important. These factors should be considered by cross-language researchers (e.g. how feasible a CLIR approach is in practice) and could form part of future evaluation tasks in campaigns such as CLEF. For example, evaluating on attributes such as how long retrieval takes, how much it costs for resources, the implications of using any “free” resources (e.g. legal constraints) and evaluating success using measures such as how many queries fail and Precision at rank 10, rather than more abstract measures such as MAP. This last aspect is important as users are often intolerant of errors and being able to correlate effectiveness of specific translation resources and approaches would help estimate their effectiveness in practice.

Given a website like the Tate Online, we propose for campaigns such as CLEF that a task which would evaluate the following aspects of multilingual access: glossaries used in the website, search (e.g. query translation), document translation of results and main static pages, and domain-specific thesauri and terminology lists, would be beneficial to a wide community.

6. CONCLUSIONS

The Tate is an important part of Britain’s cultural heritage, but increasingly recognises that its content reaches to an international audience. Like many cultural heritage institutions, the Tate sees the advantages of providing multilingual access to meet consumer demand. This paper has provided an overview of the Tate as an organisation with a real motivation for cross-language access and practical multilingual needs, looking to broaden access to its content. An important part of designing multilingual systems is to understand and define the multilingual needs of an organisation and its customers, particularly end-users of websites like Tate Online. In our work, we aim to establish the feasibility of providing multilingual access, including the role of existing translation resources in providing multilingual versions of the current website. Given the similarity of the Tate with other cultural heritage institutions, our findings and recommendations will be not be exclusive to the needs of the Tate, but also applicable to other organisations. We contend that investigating the effectiveness of multilingual search is just part of multilingual access and evaluation benchmarks such as CLEF might consider a more holistic task whereby participants must translate and provide multilingual access to a complete website (like the Tate), enabling

researchers to obtain a better estimate of the likely success of multilingual information access in practice.

Acknowledgements

Work partially supported by EU under the Information Society Technologies (IST) programme of FP6 for RTD, project MultiMATCH (Project No. 033104) and the EU FP6 project BRICKS (IST-2002-2.3.1.12). We would like to thank the Tate for their assistance in this project. In particular, we thank Jemima Rellie and Katie Dance for providing invaluable help and advice.

References

- [1] Rellie, J. (2006). 10 Years On: Hopes, Fears, Predictions and Gambles for UK Museums On-line, in J. Trant and D. Bearman (eds.). *Museums and the Web 2006: Proceedings*, Toronto: Archives and Museum Informatics.
- [2] Rellie, J. (2004). One Site Fits All: Balancing Priorities at Tate Online, in J. Trant and D. Bearman (eds.). *Museums and the Web 2004: Proceedings*, Toronto: Archives and Museum Informatics.
- [3] Parry, R., and Arbach, N. (2005). The Localized Learner: Acknowledging Distance and Situatedness in On-Line Museum Learning. In J. Trant and D. Bearman (eds.). *Museums and the Web 2005: Proceedings*, Toronto: Archives and Museum Informatics.
- [4] W3C FAQ: International and Multilingual websites, <http://www.w3.org/International/questions/qa-international-multilingual>
- [5] Multi-Lingual Web Sites: Best Practice Guidelines and Architecture (P923) Eurescom Project report, <http://www.eurescom.de/Public/projectresults/P900-series/923d1.asp>
- [6] De Troyer, O., Casteleyn, S. 2004. Designing Localized Web Sites, In *Proceedings of the 5th International Conference on Web Information Systems Engineering (WISE2004)*, pp. 547 - 558, Eds. Zhou, X., Su, S., Papazoglou, M.P., Orłowska, M.E., Jeffery, K.G., Springer-Verlag, ISBN 3-540-23894-8, Brisbane, Australia.
- [7] Del Galdo, E.M. and Nielsen, J. (Eds). 1996. *International User Interfaces*, John Wiley & Sons, New York.
- [8] Abusalah, M., Tait, J. and Oakes, M. 2005. Literature review of cross-language information retrieval, *Transactions on Engineering, Computing and Technology*, ISSN 1305-5313, 175-177.
- [9] Gonzalo, J. and Peters, C. 2005. The impact of evaluation on multilingual text retrieval. *Proceedings of SIGIR 2005*.
- [10] Hutchins, J. 2005. Computer-based translation systems and tools, State-of-the-art report for the BCS Natural Language Translation Specialist Group, http://www.bcs-mt.org.uk/nala_018.htm
- [11] Petrelli, D., Levin, S., Beaulieu, M. and Sanderson, M. 2006. Which user interaction for cross-language information retrieval? Design issues and reflections. *JASIST* 57(5): 709-722.