Supporting the Modern Polyglot - A Comparison of Multilingual Search Interfaces

Ben Steichen, Luanne Freund
University of British Columbia, Vancouver, Canada
{ben.steichen, luanne.freund}@ubc.ca

ABSTRACT

The unrelenting rise in online user diversification has generated tremendous new challenges for search system providers. Among these, the need to address multiple user language abilities and preferences is paramount. The majority of research on multilingual search has so far focused on improving retrieval and translation techniques in cross-language information retrieval. However, less research has focused on the human-computer interaction aspects of multilingual search, particularly in terms of multilingual result display interfaces. To address this research gap, this paper presents a comparison of 5 different search interface designs for multilingual search. We analyze and evaluate these interfaces through a crowd-based experiment involving 885 participants. Our results show that the common approach of interleaving multilingual results is in fact the least preferred, whereas single-page displays with clear language separation are most preferred. In addition, we show that user proficiency and search content type play an important role in user preferences, and that different interfaces elicit different user behaviors.

Author Keywords

Multilingual Search; Multilingual Interfaces; Evaluation; Human–Computer Information Retrieval; Crowd-Sourcing;

ACM Classification Keywords

H.5.m. **[Information interfaces and presentation (e.g., HCI)]**: Miscellaneous; H.3.3 **[Information Search and Retrieval]**: Search process; H.5.2 **[User Interfaces]**: Evaluation/methodology;

INTRODUCTION

According to the latest estimates by the International Telecommunication Union, there are now in excess of 2.7 billion Internet users, up from only 1 billion users in 2005.[1] The biggest contribution towards this growth comes from areas outside of the native English-speaking world, with Asian markets now accounting for over 48% of all Internet users.[2] This globalization of the online population gives rise to tremendous new challenges for adapting online services to increasingly diverse user needs, abilities, and preferences, particularly in web search.

One of the most pressing challenges lies in supporting online users' individual language skills. In particular, while the diversity of native languages among Internet users is increasing, the number of polyglot users, i.e. those who are proficient in more than one language, is also on the rise. For instance, on average 94.6% of secondary education pupils in the European Union learn English in general programs, and 64.7% learn two or more languages.[3] This growth of polyglots is equally evident throughout the world, as it is estimated that there are many more people who know English as a second language than there are native speakers.[4] This development has also directly led to new multilingual online behaviors. For instance, a recent survey showed that the majority of polyglots frequently use multiple languages during their daily online browsing and searching [16]. This is in part a result of some languages being massively underrepresented in terms of web content with respect to their user base. For example, it is estimated[5] that while 55% of all websites are in English, Chinese-language websites only account for 4%, despite 22% of Internet users being from China.[2] Given the high likelihood that the diversity of language content and the polyglot user base will continue to increase, there is a pressing need for more sophisticated information access solutions, which can address the needs of these users.

Additional challenges exist to personalize such systems to individual users and situations. For example, multilingual users have varying levels of proficiency in different languages, which may affect their interactions with systems and content. Further, the type of content being searched (e.g. general web content vs. news) may have an effect. Each of these factors may play an important role in determining what and how information should be retrieved and presented to each individual user, as suggested in [16].

Despite the trend towards Internet polyglotism, web information access systems tend to emphasize distinctions between languages, often requiring users to switch between

CHI 2015, April 18–23, 2015, Seoul, Republic of Korea.
Copyright 2015 © ACM 978-1-4503-3145-6/15/04...$15.00.
http://dx.doi.org/10.1145/2702123.2702541

[1] www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx
[2] www.internetlivestats.com/internet-users/#byregion
[3] epp.eurostat.ec.europa.eu/statistics_explained/index.php/Foreign_language_learning_statistics
[4] www.britishcouncil.org/learning-research-english-next.pdf
[5] w3techs.com/technologies/overview/content_language/all

versions of the system or conduct separate searches in order to retrieve results in more than one language. This extra effort reduces the likelihood that multilingual searches are conducted, and may result in less relevant content being found. In the long run, reliance upon such systems undermines the growing linguistic and cultural diversity of the Internet. To address these shortcomings, there has been significant progress in developing systems that can retrieve information across language barriers, i.e. cross-language search systems that allow users to search for documents in a language that is different from the original query language. However, the human-computer interaction aspects of multilingual search have received comparably less attention. In particular, there is a lack of direct comparisons of different multilingual search result displays.

The goal of this paper is to compare and evaluate the most common interface designs for multilingual search, in order to provide a better understanding of the respective user preferences and behaviors. Specifically, we aim to answer the following research questions:

- (RQ1) Which interface designs are preferred by multilingual users?
- (RQ2) Do factors such as language proficiency, type of content, or the particular languages involved influence preferences?
- (RQ3) Do different interfaces induce different user behaviors?

Through a crowd-based experiment involving 885 participants, this paper provides the first comparative analysis of commonly used multilingual search interface designs, and provides novel insights and guidelines for the design of these interfaces.

RELATED WORK

Multilingual Search Interfaces

In the field of Multilingual Search, much research has been conducted on *cross-language information retrieval*. Significant progress has been made in recent years on both translation and retrieval effectiveness, through focused campaigns such as the *Cross-Language Evaluation Forum*[6]. In comparison, the human-computer interaction aspect of such systems has received less attention, although notable exceptions include the works by Petrelli et al. [14], Oard et al [11], Marlow et al. [10], and Peters et al. [13], as well as the interactive CLEF (iCLEF) campaigns[7] (e.g. [12][8]).

Petrelli et al. [14] investigated different interaction paradigms for query elicitation and translation, by comparing a 'delegate' mode, i.e. fully automated query translation, to a 'supervised' mode, i.e. users choose/edit the translation. Through a number of studies, they showed that a compromise between the two, i.e. automated translation with user-editing capabilities, was the most effective and satisfactory approach. Similarly, studies by Oard et al. [11] found that query translation aids were generally considered advantageous to the user's interaction experience. Marlow et al. [10] investigated the effect of language ability on the use of Google Translate during multilingual search. Results showed that for unfamiliar languages, users made substantial use of automated translations, whereas for familiar languages, they tended to write their own translations and focused on webpages in the original language.

In terms of the result display, there are two commonly used approaches in multilingual information retrieval. One is to aggregate results from all languages into a single merged list. These 'Interleaved' lists can be generated using round-robin (e.g. as offered by the now removed Google Translated Foreign Pages feature), or based on collection size or relevance feedback [5]. The other approach is to use a 'Tabbed' interface, as presented in [2][8], where results are split by language and presented to users on separate tabs. Similar to these, the system in [7] allows users to switch between languages using tabs, while also affording a tab that displays *All documents* (the merging strategy for this interleaved list is not specified). Most recently, some prototypes have been developed to display results in separate 'Panels' for each language (e.g. 2lingual[8], ollito[9]).

While all these systems allow users to search across multiple languages, direct comparisons of the affordances and limitations of these designs are lacking. In this paper, we undertake this task, by conducting a comparative user study on different designs.

Monolingual Aggregate Search

While there has been relatively little research in terms of multilingual result display interfaces, several related works in the field of (monolingual) 'Aggregate Search' have investigated the general problem of displaying aggregate content from multiple source collections.

In the work by Bron et al. [6], a 'Tabbed' display is compared to a 'Blended' display in the digital library domain. In the 'Tabbed' display, results from each collection are presented separately arranged in labeled tabs, similar to 'Tabbed' multilingual search interfaces in [2][8][7]. In the 'Blended' display, results from all collections are displayed in a single list similar to the 'Universal Search' style[10] that is now commonly used by web search engines, whereby results from different search collections are shown in contiguous blocks (e.g. news results, followed by image results, followed by web

[6] CLEF - now *Conference and Labs of the Evaluation Forum* - www.clef-initiative.eu/

[7] nlp.uned.es/iCLEF/

[8] www.2lingual.com/

[9] www.ollito.com/

[10] googlepress.blogspot.ca/2007/05/google-begins-move-to-universal-search_16.html

results). User studies showed that the 'Blended' interface was particularly useful for exploring multiple sources simultaneously, while users found the 'Tabbed' display more useful to zoom in and focus on a single source.

Sushmita et al. [17] compared a 'Blended' to a 'Non-Blended' aggregate search interface in the monolingual web search domain. While the 'Blended' interface was similar to that used by Bron et al. [6], the 'Non-Blended' interface presented results from each source in separate panels similar to the abovementioned 'Panel' display. Results showed that while both interfaces had similar click frequencies, users bookmarked slightly more pages in the 'Blended' design.

Similarly, Thomas et al. [18] compared four different aggregate interfaces in the government metasearch domain. In addition to 'Tabbed', 'Blended', and 'Side-By-Side' interfaces similar to the 'Tabbed', 'Blended', and 'Panels' interfaces used in [6] and [17], they also investigated a 'More Results' interface, which contained a side-section on the main results page (i.e. the page showing results from the main source) that pointed to additional results from other sources. Their findings showed that users preferred interfaces that provided more information up-front, and that the 'More Results' interface was least preferred because no indication was given as to which other sources had relevant results.

It is important to note that unlike multilingual search interfaces, aggregated search interfaces typically focus on Image, News (which is also typically accompanied by images), and Video (e.g. [6][17]), which are much more visually salient verticals and hence easier to distinguish from web results. It is thus of interest to investigate whether findings from aggregated search interfaces would still hold in multilingual search.

Summary of Notable Interface Techniques

From the related work, we identified five commonly used techniques for displaying results from different languages or collections, namely *Tabbed*, *Interleaved*, *Panels*, *Side-Bar*, and *Universal Search*. In the context of multilingual search, in this paper:

- *Tabbed* refers to interfaces that present results from each language using separate tabs (e.g. as shown in [2][8][7][6][18]);
- *Interleaved* refers to the approach of interleaving multilingual results in a single ranked list (e.g. as in [5][7] and Google Translated Foreign Pages);
- *Panels* refers to interfaces that present results in separate panels for each language (e.g. 2Lingual, Ollito, and the 'Non-Blended' interfaces in [17][18]);
- *Side-Bar* refers to the approach of splitting 'main' and 'additional results' (e.g. as shown in the 'More Results' interface in [18]); and
- *Universal Search* refers to the approach of displaying contiguous blocks of results in different languages (e.g. the 'Blended' interfaces in [6][17][18])[11].

EXPERIMENTAL SETUP

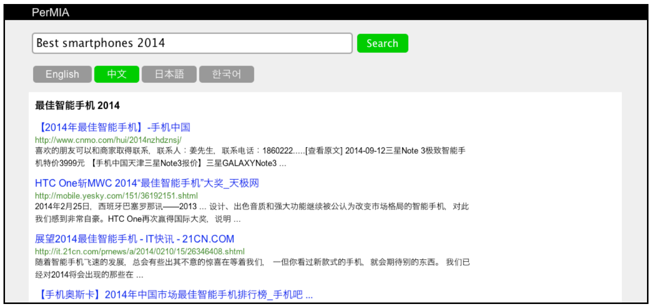
Based on the typical approaches identified in the related work, we implemented five multilingual search interfaces. These implementations [12] are representative, since they contain all characteristics of their respective interface approach discussed above.

Interfaces Used in the Study

Each of the interfaces used in the study are connected to the same back-end for web retrieval and translation (Bing Search API[13] and Microsoft Translator[14]), and only vary in terms of front-end result arrangement and interaction design.[15] These APIs are representative of the state-of-the-art in Web search and Machine Translation, which ensures a realistic experimental setup of real-world scenarios in practice, sufficient for the purpose of this study.

Tabbed Interface

As shown in Figure 1, the *Tabbed* interface provides a series of tabs allowing users to switch between their different languages, which are specified prior to the search session. The results page presents a single, monolingual ranked list, and users can choose to navigate further down/back up the list using 'next' and 'previous' buttons. The user can also choose to switch to a different language tab, at which point the query is translated and results are retrieved and displayed in that language.



最佳智能手机 2014

【2014年最佳智能手机】-手机中国
http://www.cnmo.com/hui/2014nzhdznsj/
喜欢的朋友可以和商家取得联系，联系人：姜先生，联系电话：1860222.....[查看原文] 2014-09-12三星Note 3极致智能手机特价3999元 【手机中国天津三星Note3报价】三星GALAXYNote3 ...

HTC One斩MWC 2014“最佳智能手机”大奖_天极网
http://mobile.yesky.com/151/36192151.shtml
2014年2月25日，西班牙巴塞罗那讯——2013 ... 设计、出色音质和强大功能继续被公认为改变市场格局的智能手机，对此我们感到非常自豪。HTC One再次赢得国际大奖，说明 ...

展望2014最佳智能手机 - IT快讯 - 21CN.COM
http://it.21cn.com/prnews/a/2014/0210/15/26346408.shtml
随着智能手机飞速的发展，总会有些出其不意的惊喜在等着我们，一但你看过新款式的手机，就会期待别的东西。我们已经对2014将会出现的那些在 ...

【手机奥斯卡】2014年中国市场最佳智能手机排行榜_手机吧 ...

Figure 1. Tabbed Interface

[11] We refer to this technique as *Universal Search* rather than *Blended* to better distinguish it from the *Interleaved* interface, which is technically *Blended* too.

[12] Live demo: http://www.cs.ubc.ca/~steichen/PERMIA/ - note that the feature for users to manually edit translations was not included in the study shown in this paper, as we aimed to minimize variables and distractions from a pure comparison of different layouts.

[13] datamarket.azure.com/dataset/bing/search

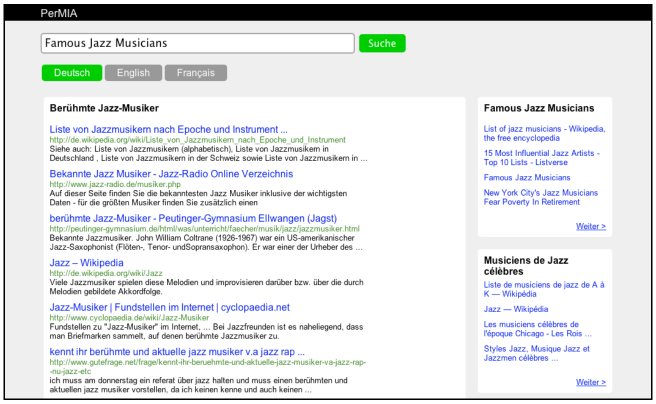
[14] datamarket.azure.com/dataset/bing/microsofttranslator

[15] When participants selected a right-to-left scripted first language (e.g. Arabic or Hebrew), the interface displayed mirrored layouts of the figures shown in this section.

In the example shown in Figure 1 the user has entered the English query 'Best smartphones 2014', and has then chosen to view Chinese results, which are retrieved using the translation'最佳智能手机 2014'.

Side-Bar Interface

As shown in Figure 2, the *Side-Bar* interface also uses tabs, and additionally contains a small side-bar section, which displays results in the user's other languages. Since the idea for this interface is to convey a split between 'main' and 'additional' results, only titles of the results are displayed in the side-bar. In the example shown, a user has chosen German ('Deutsch') as the current focus language, which displays German results in the main area on the left. On the right hand side, additional results in English and French (this user's other languages) are displayed. The top of the main result area, as well as each side bar box displays the query in the respective language.



Famous Jazz Musicians

Berühmte Jazz-Musiker

Liste von Jazzmusikern nach Epoche und Instrument ...
http://de.wikipedia.org/wiki/Liste_von_Jazzmusikern_nach_Epoche_und_Instrument
Siehe auch: Liste von Jazzmusikern (alphabetisch), Liste von Jazzmusikern in Deutschland , Liste von Jazzmusikern in der Schweiz sowie Liste von Jazzmusikern in ...

Bekannte Jazz Musiker - Jazz-Radio Online Verzeichnis
http://www.jazz-radio.de/musiker.php
Auf dieser Seite finden Sie die bekanntesten Jazz Musiker inklusive der wichtigsten Daten - für die größten Musiker finden Sie zusätzlich einen

berühmte Jazz-Musiker - Peutinger-Gymnasium Ellwangen (Jagst)
http://peutinger-gymnasium.de/html/was/unterricht/faecher/musik/jazz/jazzmusiker.html
Bekannte Jazzmusiker. John William Coltrane (1926-1967) war ein US-amerikanischer Jazz-Saxophonist (Flöten-, Tenor- undSopransaxophon). Er war einer der Urheber des ...

Jazz – Wikipedia
http://de.wikipedia.org/wiki/Jazz
Viele Jazzmusiker spielen diese Melodien und improvisieren darüber bzw. über die durch Melodien gebildete Akkordfolge.

Jazz-Musiker | Fundstellen im Internet | cyclopaedia.net
http://www.cyclopaedia.de/wiki/Jazz-Musiker
Fundstellen zu "Jazz-Musiker" im Internet, ... Bei Jazzfreunden ist es naheliegend, dass man Briefmarken sammelt, auf denen berühmte Jazzmusiker zu.

kennt ihr berühmte und aktuelle jazz musiker v.a jazz rap ...
http://www.gutefrage.net/frage/kennt-ihr-beruehmte-und-aktuelle-jazz-musiker-va-jazz-rap-nu-jazz-etc
ich muss am donnerstag ein referat über jazz halten und muss einen berühmten und aktuellen jazz musiker vorstellen, da ich keinen kenne und auch keinen ...

Famous Jazz Musicians

List of jazz musicians - Wikipedia, the free encyclopedia

15 Most Influential Jazz Artists - Top 10 Lists - Listverse

Famous Jazz Musicians

New York City's Jazz Musicians Fear Poverty In Retirement

Weiter >

Musiciens de Jazz célèbres

Liste de musiciens de jazz de A à K — Wikipédia

Jazz — Wikipédia

Les musiciens célèbres de l'époque Chicago - Les Rois ...

Styles Jazz, Musique Jazz et Jazzmen célèbres ...

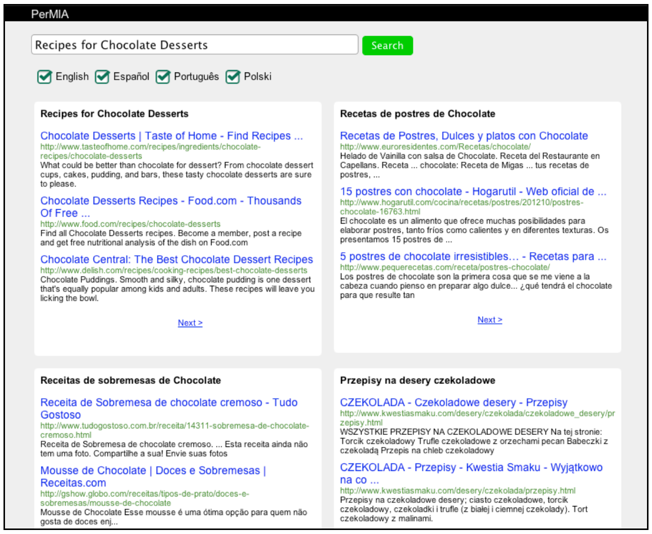
Weiter >

Figure 2. Side-Bar Interface

Panels Interface

The *Panels* interface (see Figure 3) displays results using equal-sized panels. Checkboxes allow users to select/deselect the languages that they wish to be displayed.

In order to show an overview of all languages on a single screen (i.e. above the fold), the number of results per language depends on the number of chosen languages. If a single language is chosen, the interface is similar to the *Tabbed* interface, showing a list of the top 12 results for the selected language only. When two languages are chosen, two result lists with 6 results each are displayed side-by-side. When 3 or more languages are selected, the interface displays panels that are split horizontally and vertically, with 4 results each in case of 3 languages, or 3 results each if more than 3 languages are displayed (as shown in Figure 3). This decreasing number of results per language ensures that the same number of results is displayed for each configuration. Each panel is headed by the query translated into the appropriate language, and the results within each panel can be navigated with 'previous' and 'next' buttons.



Recipes for Chocolate Desserts

☑ English ☑ Español ☑ Português ☑ Polski

Recipes for Chocolate Desserts

Chocolate Desserts | Taste of Home - Find Recipes ...
http://www.tasteofhome.com/recipes/ingredients/chocolate-recipes/chocolate-desserts
What could be better than chocolate for dessert? From chocolate dessert cups, cakes, pudding, and bars, these tasty chocolate desserts are sure to please.

Chocolate Desserts Recipes - Food.com - Thousands Of Free ...
http://www.food.com/recipes/chocolate-desserts
Find all Chocolate Desserts recipes. Become a member, post a recipe and get free nutritional analysis of the dish on Food.com

Chocolate Central: The Best Chocolate Dessert Recipes
http://www.delish.com/recipes/cooking-recipes/best-chocolate-desserts
Chocolate Puddings. Smooth and silky, chocolate pudding is one dessert that's equally popular among kids and adults. These recipes will leave you licking the bowl.

Next >

Recetas de postres de Chocolate

Recetas de Postres, Dulces y platos con Chocolate
http://www.euroresidentes.com/Recetas/chocolate/
Helado de Vainilla con salsa de Chocolate. Receta del Restaurante en Capellans. Receta ... chocolate: Receta de Migas ... tus recetas de postres, ...

15 postres con chocolate - Hogarutil - Web oficial de ...
http://www.hogarutil.com/cocina/recetas/postres/201210/postres-chocolate-16763.html
El chocolate es un alimento que ofrece muchas posibilidades para elaborar postres, tanto fríos como calientes y en diferentes texturas. Os presentamos 15 postres de ...

5 postres de chocolate irresistibles… - Recetas para ...
http://www.pequerecetas.com/receta/postres-chocolate/
Los postres de chocolate son la primera cosa que se me viene a la cabeza cuando pienso en preparar algo dulce... ¿qué tendrá el chocolate para que resulte tan

Next >

Receitas de sobremesas de Chocolate

Receita de Sobremesa de chocolate cremoso - Tudo Gostoso
http://www.tudogostoso.com.br/receita/14311-sobremesa-de-chocolate-cremoso.html
Receita de Sobremesa de chocolate cremoso. ... Esta receita ainda não tem uma foto. Compartilhe a sua! Envie suas fotos

Mousse de Chocolate | Doces e Sobremesas | Receitas.com
http://gshow.globo.com/receitas/tipos-de-prato/doces-e-sobremesas/mousse-de-chocolate
Mousse de Chocolate Esse mousse é uma ótima opção para quem não gosta de doces enj...

Przepisy na desery czekoladowe

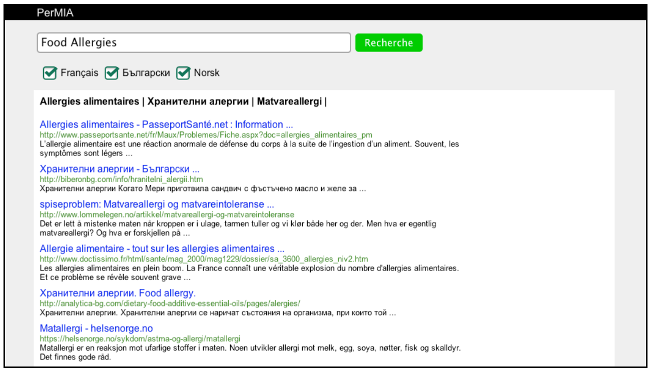
CZEKOLADA - Czekoladowe desery - Przepisy
http://www.kwestiasmaku.com/desery/czekolada/czekoladowe_desery/przepisy.html
WSZYSTKIE PRZEPISY NA CZEKOLADOWE DESERY Na tej stronie: Torcik czekoladowy Trufle czekoladowe z orzechami pecan Babeczki z czekoladą Przepis na chleb czekoladowy

CZEKOLADA - Przepisy - Kwestia Smaku - Wyjątkowo na co ...
http://www.kwestiasmaku.com/desery/czekolada/przepisy.html
Przepisy na czekoladowe desery; ciasto czekoladowe, torcik czekoladowy, czekoladki i trufle (z białej i ciemnej czekolady). Tort czekoladowy z malinami.

Figure 3. Panels Interface

Interleaved Interface

The *Interleaved* interface presents a single merged ranked list, with results from different languages being interleaved through a simple round-robin approach. For example, in Figure 4 results are shown in the following order: French-Russian-Norwegian (which is the order of language proficiency indicated by the user). In this case of 3 chosen languages, the result list displays a total of 4 results per language per page. Checkboxes allow users to select/deselect results in specific languages, and all query translations are shown together above the merged result list.



Food Allergies

☑ Français ☑ Български ☑ Norsk

Allergies alimentaires | Хранителни алергии | Matvareallergi |

Allergies alimentaires - PasseportSanté.net : Information ...
http://www.passeportsante.net/fr/Maux/Problemes/Fiche.aspx?doc=allergies_alimentaires_pm
L'allergie alimentaire est une réaction anormale de défense du corps à la suite de l'ingestion d'un aliment. Souvent, les symptômes sont légers ...

Хранителни алергии - Български ...
http://biberonbg.com/info/hranitelni_alergii.htm
Хранителни алергии Когато Мери приготвила сандвич с фъстъчено масло и желе за ...

spiseproblem: Matvareallergi og matvareintoleranse ...
http://www.lommelegen.no/artikkel/matvareallergi-og-matvareintoleranse
Det er lett å mistenke maten når kroppen er i ulage, tarmen tuller og vi klør både her og der. Men hva er egentlig matvareallergi? Og hva er forskjellen på ...

Allergie alimentaire - tout sur les allergies alimentaires ...
http://www.doctissimo.fr/html/sante/mag_2000/mag1229/dossier/sa_3600_allergies_niv2.htm
Les allergies alimentaires en plein boom. La France connaît une véritable explosion du nombre d'allergies alimentaires. Et ce problème se révèle souvent grave ...

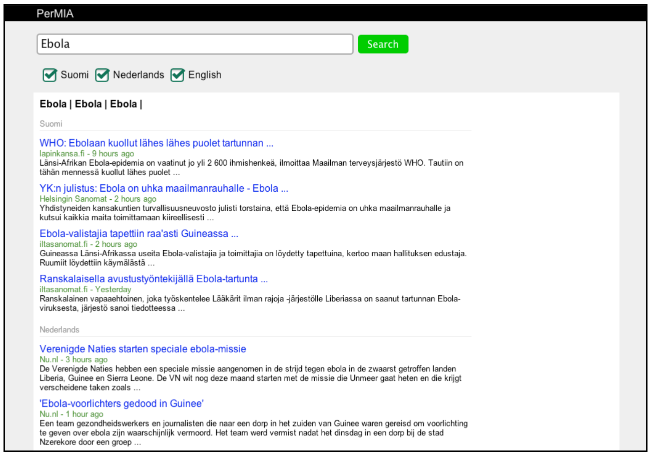
Хранителни алергии. Food allergy.
http://analytica-bg.com/dietary-food-additive-essential-oils/pages/alergies/
Хранителни алергии. Хранителни алергии се наричат състояния на организма, при които той ...

Matallergi - helsenorge.no
https://helsenorge.no/sykdom/astma-og-allergi/matallergi
Matallergi er en reaksjon mot ufarlige stoffer i maten. Noen utvikler allergi mot melk, egg, soya, nøtter, fisk og skalldyr. Det finnes gode råd.

Figure 4. Interleaved Interface

Universal Search Interface

The *Universal Search* interface (see Figure 5) also presents a single ranked list. However, rather than interleaving results, this interface presents contiguous blocks of results, in decreasing order of language proficiency. For example, in Figure 5 a batch of 4 Finnish results is followed by a batch of 4 Dutch results, etc. Checkboxes allow users to select/deselect their languages, and all query translations are shown at the top of the list.



Ebola | Ebola | Ebola |

Suomi

WHO: Ebolaan kuollut lähes lähes puolet tartunnan ...
lapinkansa.fi - 9 hours ago
Länsi-Afrikan Ebola-epidemia on vaatinut jo yli 2 600 ihmishenkeä, ilmoittaa Maailman terveysjärjestö WHO. Tautiin on tähän mennessä kuollut lähes puolet ...

YK:n julistus: Ebola on uhka maailmanrauhalle - Ebola ...
Helsingin Sanomat - 2 hours ago
Yhdistyneiden kansakuntien turvallisuusneuvosto julisti torstaina, että Ebola-epidemia on uhka maailmanrauhalle ja kutsui kaikkia maita toimittamaan kiireellisesti ...

Ebola-valistajia tapettiin raa'asti Guineassa ...
iltasanomat.fi - 2 hours ago
Guineassa Länsi-Afrikassa useita Ebola-valistajia ja toimittajia on löydetty tapettuina, kertoo maan hallituksen edustaja. Ruumiit löydettiin käymälästä ...

Ranskalaisella avustustyöntekijällä Ebola-tartunta ...
iltasanomat.fi - Yesterday
Ranskalainen vapaaehtoinen, joka työskentelee Lääkärit ilman rajoja -järjestölle Liberiassa on saanut tartunnan Ebola-viruksesta, järjestö sanoi tiedotteessa ...

Nederlands

Verenigde Naties starten speciale ebola-missie
Nu.nl - 3 hours ago
De Verenigde Naties hebben een speciale missie aangenomen in de strijd tegen ebola in de zwaarst getroffen landen Liberia, Guinee en Sierra Leone. De VN wit nog deze maand starten met de missie die Unmeer gaat heten en die krijgt verscheidene taken zoals ...

'Ebola-voorlichters gedood in Guinee'
Nu.nl - 1 hour ago
Een team gezondheidswerkers en journalisten die naar een dorp in het zuiden van Guinee waren gereisd om voorlichting te geven over ebola zijn waarschijnlijk vermoord. Het team werd vermist nadat het dinsdag in een dorp bij de stad Nzerekore door een groep ...

Figure 5. Universal Search Interface

Participant Recruitment

We conducted a crowd-based user study using the Crowdflower[16] platform. Participants were required to have 'some proficiency'[17] in at least two out of a set of 32 languages (which are all the languages supported by the Microsoft Bing and Translator APIs). In addition, only participants using desktop or laptop computers were recruited (future studies may specifically focus on mobile devices). To reduce noise and increase quality, only Tier 3 contributors [18] were invited for participation. Each participant was paid 10 cents, and the average user completion time was approximately 6 minutes.

Search Queries

Since the goal of our study was to compare different results display designs, it was necessary to ensure that while layouts varied, all other variables remained fixed, such as the search queries and their translations. In particular, we did not want preferences for a particular interface to be influenced by result relevance or better query translations. We therefore controlled the queries and manually verified query translations. This setup thus ensured that participants were presented with the same content, with the only variable being the layout of this content. To provide an unbiased and diverse set of topics, our queries were taken from the CLEF 2008 Ad Hoc Track [1] and Google Trends[19], as shown in Table 1.

To test for the effect of different content types (i.e. to answer RQ2), some queries were used to retrieve general Web content, whereas other queries (regarding current events) were used to retrieve News content[20]. This was based on our hypothesis that these broad distinctions might elicit different preferences (e.g. News searches particularly favoring displays that provide an overview of diversified viewpoints).[21] To avoid source mixing, each query returned results from one source only - either Web or News.

Query	Type
Decorating Children's Rooms	Web
Recipes for Chocolate Desserts	Web
Bordeaux Wine Guides	Web
Famous Jazz Musicians	Web
Youth Employment in Europe	Web
Food Allergies	Web
Diabetes	Web
Best smartphones 2014	Web
Ukraine	News
Gaza	News
Ebola	News
FIFA World Cup	News

Table 1. Queries used for the study

Procedure

After giving their informed consent, participants were first asked to provide demographic information regarding their country of origin and country of residence, as well as the languages in which they had some proficiency. Participants were also asked to indicate their levels of proficiency in the selected languages on a 5-point scale with labeled ends, where 1 indicates 'elementary proficiency' and 5 indicates 'native or bilingual proficiency'.

Participants were then shown a pair-wise comparison of results for one of the queries using two of the five interfaces (counterbalanced across participants overall) embedded in the Crowdflower interface.[22] Participants were free to click on any of the results, or use any other functionality (e.g. switch language using buttons, see more results, etc.), although the search box and the search button were disabled (to reduce variables and minimize distractions as discussed previously). The search boxes in the interfaces were pre-filled with the same query, drawn from the queries shown in Table 1. We asked users to indicate what they found to be 'the most interesting' result in each of the interfaces. Users were then asked to indicate which interface they preferred (or No Preference) and why (open question).

As noted in [9], first impressions or short interactions are crucial in human decision-making and preference judgments. Our study thus focused on relatively short interactions to gauge user interface preferences, as opposed to evaluating lengthy interactions with a fully operational multilingual search system. This setup minimizes

[16] www.crowdflower.com

[17] Having 'some proficiency' was defined as 'you must have some reading/writing ability' in the language.

[18] Described as '… the highest performance contributors who account for 7% of monthly judgments and maintain the highest level of accuracy …' – www.crowdflower.com

[19] www.google.com/trends/

[20] Using the 'source' property in the Bing Search API

[21] Future studies may also investigate other, more fine-grained topic distinctions (e.g. health, technology, etc.)

[22] Following the technical approach of [20].

distractions posed by operational factors, such as the quality of the ranking algorithm or the query translations.

As a quality control mechanism, users were asked to indicate what the topic of the search was, by selecting one correct answer out of five topics from a pre-existing list. We also asked users to describe the search topic in the languages they claimed to speak. We then manually verified these responses and discarded spam entries. Although noise is unlikely to be completely eliminated due to the nature of self-reported data, we believe the control exercised in this study is sufficient for quality assurance.

Participant responses to the demographic questions and their interface preferences were collected through the Crowdflower interface. In addition, all interactions with the interface were logged by the search system, including language choices, checkboxes/tabs, and result clicks.

Data Analysis

We used the Bradley-Terry model [4] for the statistical analysis of preferences. Given a number of independently sampled pairwise comparisons (in our case preference judgments between pairs of interfaces), the Bradley-Terry model generates an overall 'ability estimate' for each item using transitivity, which can be interpreted as a generalized overall 'interface preference score'. These values not only allow an overall ranking of the different interfaces, but also provide odds ratios of preferences between any pair of items, which can then be converted to a more interpretable value that indicates the 'probability that one interface is preferred over the other' (thereby answering RQ1). For this analysis, we used the BradleyTerry2 package for R [19], which generates overall ability scores, as well as statistical significance testing between the 'ability' of pairs of items.

Note that a Bradley-Terry model allows for the specification of model factors to examine the effect and its statistical significance of study factors (analogous to factors in standard regression or ANCOVA). In order to answer RQ2, we specified proficiency ratings, the content type (Web vs. News), and a user's first language as model factors. For proficiency, we first ran a model using the actual value, followed by a 'high' vs 'low' model (using a median split) for ease of presentation. All reported results are corrected using Bonferroni corrections for Type I errors.

To analyze users' behaviors (to answer RQ3), we ran chi-square tests on result clicks, as well as choices of 'most interesting results', again correcting pairwise comparisons using Bonferroni corrections.

RESULTS

A total of 1220 participants took part in the study, of which responses were retained from 885 after filtering out invalid entries using the quality control mechanism discussed above. In the following sections, we provide an overview of participant demographics, followed by an analysis of user preferences and user behaviors.

Participant Demographics

A total of 76 countries of origin (see Figure 6), and 69 countries of current residence were reported by our participants. As required by the study, all users were proficient in at least two languages. In addition, 193 participants (21.8%) indicated that they were proficient in three languages, 48 participants were proficient in four languages (5.4%), and 9 participants (1%) indicated they were proficient in more than four languages.

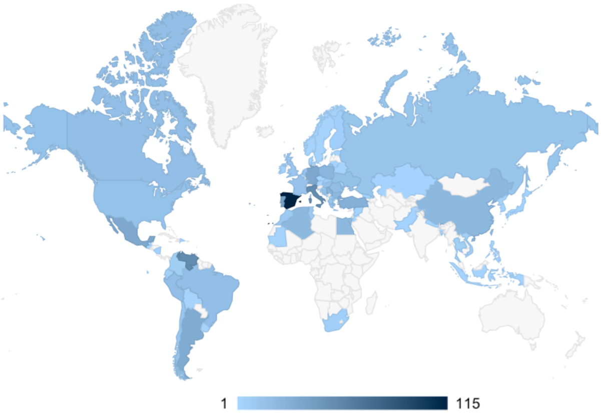


Figure 6. Map of participants' countries of origin - color tones reflect the number of participants from each country

In terms of users' first language (i.e. their most proficient language - L1 hereafter), 31 different languages were reported, with the most common languages being Spanish (32.1%), English (13.6%), Arabic (6.4%), Portuguese (6.1%), Italian (6.0%), and Chinese (5.1%)[23] (see Figure 7).

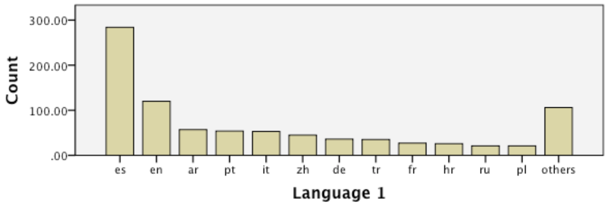


Figure 7. Overview of participants' L1

For L2 (i.e. their second most proficient language), 25 different languages were reported, with the vast majority of participants (71.2%) indicating English as their second language, followed by French (8.4%), and Spanish (4%) (see Figure 8).

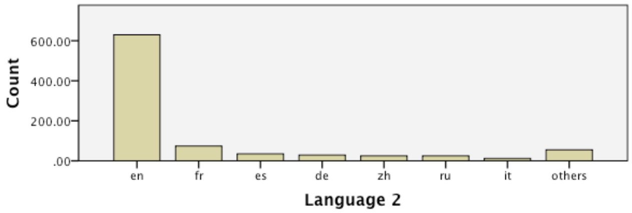


Figure 8. Overview of participants' L2

[23] note that we merged results for 'Chinese Simplified' and 'Chinese Traditional'

For people who indicated 3 or more languages, the most common language for 'L3' was English (33.6%), followed by French (18.8%) and German (12.8). For 'L4', the most common languages were French (26.3%), Spanish (19.3%), and Italian (10.5%).

Since L1 was participants' strongest language, the proficiency level for this was most dominantly 5 (i.e. native or bilingual language – 89.8%) or 4 (7.6%). For L2, there were great variations in terms of proficiencies, as shown in Figure 9. Lastly, users' third and fourth language proficiency levels were predominantly 3 or lower.

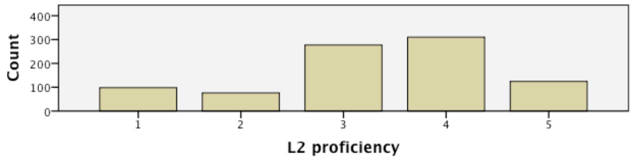


Figure 9. Overview of L2 proficiency levels (1=elementary proficiency, 5=native or bilingual proficiency)

Results for User Preferences (RQ1)

Recall that the primary research question addressed in the study is to find out what type of multilingual search result display is most preferred by multilingual users. To answer this question, we ran a Bradley-Terry test using the set of 885 preference judgments. As shown in Figure 10, there are indeed significant differences between interfaces, with the *Panels* interface having the highest overall 'preference score' (i.e. highest 'ability estimate'), and the *Interleaved* interface having the lowest.

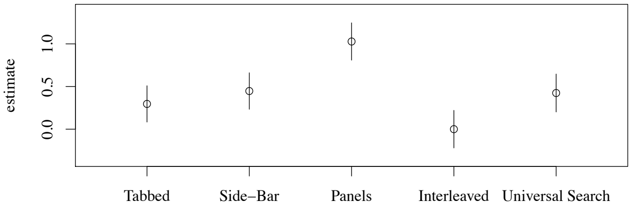


Figure 10. Overall comparison of the 5 interfaces. Ability estimates are generated from the Bradley Terry model, illustrating overall preferences for each interface

We found that the preference for *Panels* is statistically significantly higher than all other interfaces (p<.001), and the preference for *Interleaved* is statistically significantly lower than all interfaces (p<.001), except *Tabbed*.

	Tabbed	Side-Bar	Panels	Inter-leaved	Univ. Search
Tabbed		46.23%	32.48%	57.35%	46.81%
Side-Bar	53.77%		35.88%	**62.00%**	50.58%
Panels	**67.52%**	**64.12%**		**73.65%**	**64.66%**
Interleaved	42.65%	38.00%	26.35%		38.56%
Univ. Search	53.19%	49.42%	35.34%	**61.44%**	

Table 2. Direct comparisons between interfaces (generated using Bradley-Terry ability estimates) – bold numbers indicate that the interface in the left column is statistically significantly preferred over the corresponding interface in the top row

As shown in Table 2, the probability of choosing the *Panels* interface over the *Interleaved* interface is 73.65%. Similarly, the probability of choosing the *Universal Search* interface over the *Interleaved* interface is 61.44%.

Influences of Proficiency, Language, and Content Type on User Preferences (RQ2)

First, we tested for an effect of level of proficiency on interface preference. In this analysis, we found that there was indeed a statistically significant effect (p<.001) of a user's proficiency level in their L2.[24]

When splitting participants into either having a 'high' or 'low' proficiency in their second language (defined as proficiency levels of 5/4 and 3/2/1 respectively, based on a median split), we found that users with high proficiencies in their L2 significantly preferred the *Universal Search* interface over the *Interleaved* interface, and preferred it almost to the same extent as the *Panels* interface (see Figure 11 top figure). By contrast, for users with a low proficiency in their L2 (Figure 11 bottom figure), the *Universal Search* interface is no longer better than the *Interleaved* interface.

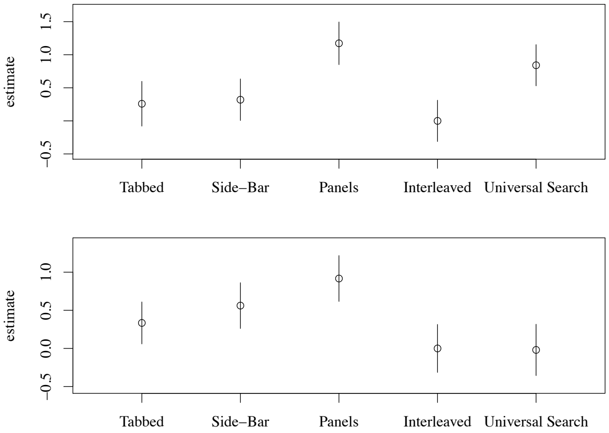


Figure 11. Interface preferences for participants with 'high' (top) and 'low' (bottom) proficiencies in their L2.

Secondly, we tested for an effect of the particular languages in which users were proficient, especially in terms of first languages. While we did not find a statistically significant effect, there were several trends suggesting that the user's first language may have some influence on interface preferences.[25] In particular, we found that users with right-to-left scripted first languages *highly* preferred the *Panels* interface, with all other interfaces being equally

[24] The majority of participants who were proficient in a third/fourth language indicated low proficiencies in them. Since there is little variation in L3 and L4 proficiencies, we did not perform separate tests for these.

[25] which may also be related to general cultural differences, as discussed in [15].

dispreferred. This may be a result of the majority of these users receiving both right-to-left (i.e. Arabic or Hebrew) and left-to-right (e.g. English) results, suggesting that the *Panels* interface provides better support for such 'mixed' script direction scenarios. Given the low number of participants for many language groups, however, there may have been insufficient power to show any statistically significant differences. While this study was explicitly aimed at gathering overall preferences from a diverse range of participants, a future study could focus on a few select languages to better elicit such effects. Similarly, we did not find an effect of the number of languages in which a user was proficient (e.g. two, three, four).

Lastly, we analyzed whether differences in content types influenced interface preferences. Results showed a statistically significant difference between Web and News content (p<.01). In particular, for News search, the preference for *Side-Bar* increased significantly, becoming statistically significantly preferred over the *Tabbed* and *Universal Search* interfaces (in addition to being preferred over the *Interleaved*). By contrast, for Web searches, the *Side-Bar* Interface was not preferred over *Tabbed*, *Universal Search*, or *Interleaved*.

User Behavior (RQ3)

Recall we were interested in whether different interfaces would lead to i) different choices (in terms of language) for the 'most interesting result', and ii) different click behaviors. Note that we do not assume more/less choices or clicks in a particular language to be more/less desirable or optimal. The goal here is to simply compare behaviors induced by different interfaces, in order to gain a better understanding of the different designs.

'Most interesting result' analysis

We first ran chi-square tests to analyze whether variation in interface design was associated with differences in the language of results chosen as 'most interesting', i.e. whether they were participants' L1, L2, etc. We found a significant difference between interfaces (p<.001), with participants choosing more results in their L1 when using the *Tabbed* and *Side-Bar* interfaces (see Figure 12). Conversely, the *Panels*, *Interleaved*, and *Universal Search* interfaces were associated with a significantly more 'diversified' selection of results in terms of languages.

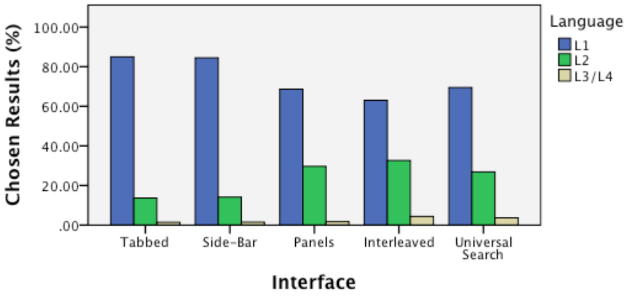


Figure 12. 'Most interesting result' selections by language (L1 vs. L2 vs. L3/L4)

Result clicks

The analysis of clicks (i.e. users clicking on a search result) showed a very similar pattern, with statistically significant differences between interfaces (p<.001). In particular, very few clicks were performed in a language other than participants' first language in both the *Tabbed* and *Side-Bar* interfaces (see Figure 13). Conversely, the *Panels*, *Interleaved*, and *Universal Search* interfaces led users to many more clicks in their other languages.

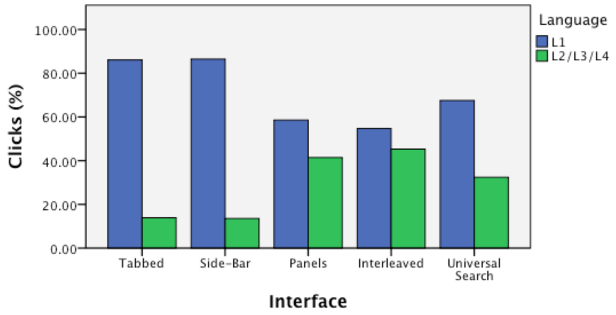


Figure 13. Result clicks by language (L1 vs. L2/L3/L4)

SUMMARY & DISCUSSION

The analysis of users' preferences, result choices, and click behaviors has shown that there are significant differences between different multilingual search interface designs. In this section, we will summarize key findings, provide further insights and recommendations based on participants' open-text comments, and discuss implications for these designs.

The Panels interface was most preferred over all other interfaces, while the Interleaved interface, most commonly used in multilingual search applications, was least preferred. Many participants considered the search results in the *Panels* interface to be more organized, easier to navigate, and easier to read. In contrast, users mentioned that the *Interleaved* interface was confusing, and that the unorganized languages made it harder to navigate and even tiring. These findings suggest that multilingual search interfaces should be designed to provide clear divisions between results in different languages. In particular, the preference for the *Panels* design is much stronger than previous research findings in monolingual aggregate search. This suggests that the inherent saliency differences between traditional verticals (e.g. between web and image results) help users in the separation of results, whereas multilingual results require additional structuring. Indeed, a possible reason for the negative results and comments on the *Interleaved* interface may be that users frequently needed to 'code-switch' between languages without any structure, nor language indication, which may have increased the cognitive load. A potential solution for this may be to use labels, icons, colors, or flags (although flags are generally regarded as bad practice in the localization community), which can be explored in future studies.

The degree of separation afforded in a Tabbed design seems to be too extreme - participants criticized the *Tabbed* interface for not showing what other results were available, and requiring too much clicking. This finding suggests that the *Tabbed* interface provides insufficient context to users and requires extra effort in switching between languages. These results are consistent with previous studies on monolingual aggregated search interfaces [18], which suggested that aggregate interfaces should show information about available information 'up-front' as much as possible. The fact that the *Tabbed* interface is dispreferred to all other interfaces (except *Interleaved*, of which the specific deficiencies are discussed above) also suggests that multilingual users are, in general, in favor of interfaces that show multiple languages on a single screen. This is possibly due to the increased information breadth of the other interfaces, as participants found them more varied, diverse, and providing a more complete overview.

Universal Search is the preferred design if only a single list is to be presented (e.g. to avoid drastically changing modern web search engine interfaces), particularly for users with high second language proficiency. These participants identified as positive features the simplicity of this design and the clear separation of languages in a single list. However, participants with low proficiency in their second languages dispreferred both single-list designs, i.e. *Universal Search* and *Interleaved*, to all others. This is potentially caused by the degree of language separation being insufficient for such users. This suggests that the *Universal Search* interface is well-suited only to users with high second language proficiencies, and that single-list designs should be avoided altogether for users with low second language proficiencies.

Preferences for the Side-Bar interface were similar to those of the Tabbed interface, even though *Side-Bar* provided additional information in other languages in a manner similar to the *Panels* interface. A likely cause for this result is that many participants mistook the side-bar for an advertising area, and they consciously chose to ignore these results, hence judging this interface to be less useful. While this seems to be the case for the Web searches, for News searches, however, the *Side-Bar* interface was actually preferred over the *Tabbed* interface, and came close to the *Panels* interface in terms of preference score. Participants mentioned that this interface was clean, and allowed them to focus on one language while being aware of/not missing others. This finding suggests that the *Side-Bar* design may be more effective for particular types of search tasks associated with news search, such as surveying and comparing multiple perspectives on a particular news story. Furthermore, it is possible that the performance of the *Side-Bar* interface may be improved significantly by making it clear to users that the side-bar results are in fact additional results in other languages (e.g. through explicit labeling at the top of the side-bar, or by moving the side-bar to a different location altogether). This may make the *Side-Bar* a viable option for quickly adding multilingual features to modern web search engine interfaces.

Many participants chose to select results in a language other than their first language as the 'most interesting result' – this result confirms the findings from [16] that polyglot users frequently acquire information in different languages. In particular, participants commented positively on the inclusion of multilingual functionalities regardless of the interface, saying there are more choices overall - not just results from one country - and that sometimes the results are more relevant and current in one language than in others. Although our results have shown that users often click on results in non-primary languages, it may not be the case that every user actually benefits from diversified multilingual results (e.g. for a simple fact finding task, multilingual results may not be necessary, and can be detrimental to the search process under some circumstances). Depending on the goal of the system designer and the user's information needs, this aspect of stimulating certain user behaviors may be explored in future studies that specifically focus on task and query types.

Lastly, *interface designs influence user behaviors* - we found that given different interfaces, more/less result clicks and choices in non-primary languages were generated. This difference may be influenced by the perceived ranking position, which is an unavoidable artifact intrinsically associated with all interface designs. For example, the low diversity for the *Tabbed* interface is intuitive, since it requires a user interaction (language tab switching) to view results in languages other than L1. For *Side-Bar*, the low diversity may indicate that users ignored the extra results (mistaking them for advertisements). Meanwhile, the greatest diversity of selected results was obtained using the interface designs preferred most (*Panels*) and least (*Interleaved*). This finding suggests that diversity is valued when the user is aware of and feeling in control of it, as in the *Panels* design, but not when it is imposed upon the user by the system. Although we do not make any assumptions about more/less clicking in (non-)primary languages being an advantage/disadvantage, our findings, however, may be useful to inform those aiming to build systems that maximize/minimize click diversity if so desired.

CONCLUSION & FUTURE WORK

This paper presents the first direct comparison of multilingual search result display designs with several informative key findings and recommendations (discussed above), thereby providing a significant contribution to the development, improvement, and deployment of systems that aim to better support the ever-increasing number of global polyglot users.

Using the most common designs in both multilingual information retrieval and monolingual aggregate search, our crowd-based evaluation has shown that participants particularly like an interface that presents results from different languages in separate 'Panels' on the same screen, and that language separation is generally preferred over the common approach of interleaving results. Also, our results have shown that the user's proficiency and search content type play a role in user preferences, and that different interfaces elicit different user behaviors.

In future work, we intend to conduct studies that specifically focus on different types of search tasks (e.g. fact finding, information gathering) and topics (e.g. business, health, technology) to determine which benefit most from linguistic diversity in search results, as well as which languages are preferred for which tasks and topics. In a more *personalized* multilingual search system (as proposed in [16]), user preferences and contextual factors could then be integrated into each of the interfaces, for example, through a reorganization of panels in the *Panels* interface, by slotting the most relevant languages higher up in the *Universal Search* interface, or by dynamically selecting the number of results per language.

Lastly, while the specific focus of this paper was to elicit user preferences without potential distractions posed by an operational interactive multilingual search system, such as the quality of the ranking algorithm or the query translations, it can be argued that some issues (e.g. confusing the side-bar with advertisements) may be overcome if given user familiarity. This speculation can be verified in future studies, e.g. through the evaluation of a fully functional interactive retrieval system involving simulated work tasks [3] that require users to enter their own queries.

ACKNOWLEDGEMENTS

This research is supported by the National Research Fund Luxembourg and cofunded under the Marie Curie Actions of the European Commission (FP7-COFUND).

REFERENCES

1. Agirre, E., Nunzio, G.M.D., Ferro, N., Mandl, T., and Peters, C. CLEF 2008: Ad Hoc Track Overview. *Evaluating Systems for Multilingual and Multimodal Information Access*, (2009), 15–37.
2. Amato, G., Debole, F., Peters, C., and Savino, P. The MultiMatch Prototype: Multilingual/Multimedia Search for Cultural Heritage Objects. *Research and Adv. Technol. for Digital Libraries*, (2008), 385–387.
3. Borlund, P. The IIR evaluation model: a framework for evaluation of interactive information retrieval systems. *Information Research 8*, 3 (2003).
4. Bradley, R.A. Paired comparisons: Some basic procedures and examples. *Nonparametric Methods 4*, (1984), 299–326.
5. Braschler, M. Combination Approaches for Multilingual Text Retrieval. *Information Retrieval*, *7*, 1-2 (2004), 183–204.
6. Bron, M., van Gorp, J., Nack, F., Baltussen, L.B., and de Rijke, M. Aggregated Search Interface Preferences in Multi-session Search Tasks. *Proc. of 36th Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, (2013), 123–132.
7. Capstick, J., Diagne, A.K., Erbach, G., Uszkoreit, H., Leisenberg, A., and Leisenberg, M. A system for supporting cross-lingual information retrieval. *Inf. Process. & Manage. 36*, 2 (2000), 275–289.
8. Clough, P., Al-Maskari, A., and Darwish, K. Providing Multilingual Access to FLICKR for Arabic Users. *Evaluation of Multilingual and Multi-modal Information Retrieval*, (2007), 205–216.
9. Gladwell, M. *Blink: The Power of Thinking Without Thinking.* Back Bay Books, New York, 2006.
10. Marlow, J., Clough, P., Recuero, J.C., and Artiles, J. Exploring the effects of language skills on multilingual web search. *Proc. of 30th European conf. on Advances in information retrieval*, (2008), 126–137.
11. Oard, D.W., He, D., and Wang, J. User-assisted query translation for interactive cross-language information retrieval. *Inf. Proc. & Manage. 44*, 1 (2008), 181–211.
12. Peinado, V., Artiles, J., Barker, E., and Lopez-Ostenero, F. FlickLing: a Multilingual Search Interface for Flickr. *Working Notes - CLEF Workshop*, (2008).
13. Peters, C., Braschler, M., and Clough, P. *Multilingual Information Retrieval: From Research To Practice, chapt. 4,* Springer Science & Business Media, (2012).
14. Petrelli, D., Levin, S., Beaulieu, M., and Sanderson, M. Which user interaction for cross-language information retrieval? Design issues and reflections. *J. Assoc. Inf. Sci. Technol., 57*, 5 (2006), 709–722.
15. Reinecke, K. and Gajos, K.Z. Quantifying Visual Preferences Around the World. *Proc. of SIGCHI Conf. on Human Factors in Computing Syst.*, (2014), 11–20.
16. Steichen, B., Ghorab, M.R., Lawless, S., O'Connor, A., and Wade, V. Towards Personalized Multilingual Information Access - Exploring the Browsing and Search Behavior of Multilingual Users. *Proc. of int. conf. on User Modeling, Adaptation, and Personalization*, (2014), 435-446.
17. Sushmita, S., Joho, H., Lalmas, M., and Villa, R. Factors Affecting Click-through Behavior in Aggregated Search Interfaces. *Proc. of 19th Int. Conf. on Inf. and Knowledge Management*, (2010), 519–528.
18. Thomas, P., Noack, K., and Paris, C. Evaluating Interfaces for Government Metasearch. *Proc. of 3rd Symp. on Inform. Interac. in Context*, (2010), 65–74.
19. Turner, H. and Firth, D. Bradley-Terry models in R: the BradleyTerry2 package. *Journal of Statistical Software 48*, 10 (2012).
20. Zuccon, G., Leelanupab, T., Whiting, S., Yilmaz, E., Jose, J.M., and Azzopardi, L. Crowdsourcing interactions: using crowdsourcing for evaluating interactive information retrieval systems. *Information Retrieval 16*, 2 (2013), 267–305.