BantuWeb: A Digital Library for Resource Scarce South African Languages

Andreas von Holy
University of Cape Town
Private Bag X3
Cape Town, South Africa 7701
vhland002@myuct.ac.za

Alon Bresler
University of Cape Town
Private Bag X3
Cape Town, South Africa 7701
brsalo001@myuct.ac.za

Osher Shuman
University of Cape Town
Private Bag X3
Cape Town, South Africa 7701
shmosh001@myuct.ac.za

Catherine Chavula
University of Cape Town
Private Bag X3
Cape Town, South Africa 7701
cchavula@cs.uct.ac.za

Hussein Suleman
University of Cape Town
Private Bag X3
Cape Town, South Africa 7701
hussein@cs.uct.ac.za

ABSTRACT

South Africa is a linguistically diverse country: it is a home to 11 official languages of which nine, excluding English and Afrikaans, are Resource Scarce Languages (RSLs). Accordingly, many South Africans struggle to access information written in their native languages on the Web. Unfortunately, lack of access to information hinders social economic growth. This paper proposes a Web based digital library to act as a central repository for content written in these languages that is crawled from the Web, and generated or contributed by a community of users. Gamification features have been incorporated into the digital library to motivate users to contribute content to strengthen the collection of resources and to increase community participation. Specifically, the paper: (i) proposes a ranking algorithm, smart interleaving, to aggregate and rank multilingual search results effectively from collections of varying size; and (ii) investigates which gamification features, among leaderboard, notifications, virtual points and level, motivate users to contribute content in the context of South African RSLs. The results show that users were motivated to contribute more content to reach the next level than improving their leaderboard ranking or virtual points. Further, the overall results on merging and ranking multilingual search results show no significant improvement in using smart interleaving.

CCS CONCEPTS

•Information systems → Digital libraries and archives; Multilingual and cross-lingual retrieval; *Presentation of retrieval results;* Rank aggregation; Distributed retrieval;

SAICSIT '17, Thaba Nchu, South Africa

978-1-4503-5250-5/17/09…$15.00
DOI: 10.1145/3129416.3129446

KEYWORDS

Digital Libraries, Gamification, Crowdsourcing, Multilingual Information Retrieval, Search Engines, Information Retrieval Evaluation, Web Crawling, Language Preservation

ACM Reference format:
Andreas von Holy, Alon Bresler, Osher Shuman, Catherine Chavula, and Hussein Suleman. 2017. BantuWeb: A Digital Library for Resource Scarce South African Languages. In *Proceedings of SAICSIT '17, Thaba Nchu, South Africa, September 26–28, 2017,* 10 pages.
DOI: 10.1145/3129416.3129446

1 INTRODUCTION

South Africa is a multilingual country with 11 official languages: Afrikaans, English, isiNdebele, isiXhosa, isiZulu, Sepedi, Sesotho, Setswana, siSwati, Tshivenda and Xitsonga [5, 23]. Nine of the languages, excluding English and Afrikaans, have minimal content available online [19], labeling these languages as Resource Scarce Languages (RSLs). Afrikaans and English are well documented and researched [11, 19, 24]. Although, English is the lingua franca, the majority of the population speaks at least one of the nine official African languages [23]. Limited resources on-line reduces the chance of the culture of the language speakers and the language itself to spread.

This paper proposes a centralized online digital library – BantuWeb – that features Web pages and digital documents in South African RSLs. BantuWeb aims to make two contributions: firstly, tools for finding digital content and creating content in South African RSLs; and, secondly, insights in motivating speakers of South African RSLs to create content in these languages. BantuWeb contains content harvested from the Web and submitted by users. The portal aims to create a sense of excitement around RSL preservation as well as to create a rich sub-Web of content that grows over time, referred to as an 'organic corpus' [8]. Gamification elements are incorporated into the portal's features to motivate users to contribute their own content.

Gamification refers to the use of elements usually found in games, being used in systems where they would usually not be found [10]. The use of gamification elements aims to improve user experience of the system [10, 16]. A leaderboard, virtual points, levels and

notifications are the gamification features used in Bantuweb. Gamification has shown to produce positive effects such as increasing user interaction and service usage. However, many studies have shown that this effect is context dependent and user type oriented [16]. Accordingly, the effect of the gamified features was studied to understand how users of Bantuweb would interact with and experience gamification. Moreover, previous studies have not explored gamification on a digital library of RSLs.

BantuWeb consists of three major components: (i) a language focused crawler, which crawls and indexes Web pages written in the nine languages; (ii) a Web portal, which is a central location for users to access and contribute resources to increase the presence of their language on the Internet; and (iii) a multilingual search feature, which allows users to search for information in languages of their choice. Supporting the search experience of polyglots is a complex task: multilingual results need to be presented in a manner that is useful to the context of the user [7]. User search experience and evaluation of multilingual search results specifically in the context of the digital library are studied.

The remainder of the paper is structured as follows. Section 2 discuses related work. Section 3 provides the system design overview. Section 4 discusses experimental design and results. Section 5 discusses the results and Section 6 concludes.

2 RELATED WORK

The proceeding sections discuss features used in gamification and their effect on the experience of users, and Multilingual Information Retrieval (MLIR) search results aggregation and ranking algorithms and strategies that have been proposed in literature.

2.1 Gamification

Gamification is an umbrella term for various features taken from games that have been applied in non-game contexts. Gamification techniques have been applied in various fields, primarily in education, intra-organisational systems and work environments. A literature study by Hamari *et al.* [16] compared 24 empirical peer-reviewed papers on gamification, and they found that leaderboards, virtual points and achievements were the most used gamification features. Leaderboards rely on a ranking system that is usually implemented with virtual points or number of achievements. This co-dependence is a strong suggestion why these three features are favourites among gamification features.

2.1.1 Achievements. Achievement gamification systems use a rewarding strategy that gives extra goals to users for them to be granted more points. Denny [9] made use of achievements in an educational context, where students completed achievements by contributing to an online learning tool. Denny's study [9] reported positive user feedback and improved quality of contributions throughout the time of the study. Montola *et al.* [22] used achievements to motivate users of a photograph sharing application and the majority of users reported an improved user experience. However, they found that users fit into three groups: (i) users who enjoyed the features; (ii) users who disliked features; and (iii) users who did not understand the purpose of the features [22]. This study shows that gamification features can sometimes improve user experience of a subset of users.

2.1.2 Virtual Points. Virtual points are rewarded to users to perform a specific task on a gamified system. Farzan *et al.* [12] investigated user motivation using virtual points on a social networking site in two studies [12, 13]. The follow up study [12] involved deploying a system used in the pilot study [13], to a bigger user base and over a longer period of time. These are two objectives that are part of BantuWeb's future work. Farzan *et al.* [12, 13] found that points are an effective motivation technique in both their evaluations. The follow-up study [12] had an opt-out option from the gamification features, similar to Montola *et al.* [22]. This is because some users prefer to keep their contributions and actions private and disliked the publicity of their actions introduced through gamification features [12]. Moreover, some users may dislike gamification features and systems should be available without gamification [12, 22].

2.1.3 Leaderboards. Leaderboards in combination with virtual points are the most used and successful gamification features. Packham and Suleman [26] used leaderboards and virtual points in a crowdsourcing application to translate text from English to isiXhosa. In another study, Havenga *et al.* [17] created a social media platform for sharing heritage photographs and incorporated leaderboards and badges to motivate contributions. Havenga *et al.* [17] revealed that the leaderboard feature had a greater effect than badges. An important observation made by Havenga *et al.* [17] is that a lot of valuable content was gathered through the contributions made by the users that could not have been collected otherwise.

Packham and Suleman [26] explored the effect of payments on user participation or contribution, and found that participants were reluctant to contribute without payment. Further, Packham and Suleman [26] proposed that participants were motivated more by fixed payment rates rather than performance based rewards. The study conducted on BantuWeb included an evaluation on what factors motivated the participants to contribute content.

2.2 Merging Algorithms for Multilingual Search Results

A multilingual information retrieval system returns search results in multiple languages in response to a query submitted by a user in a single language [27]. Current MLIR systems are modelled after federated search systems, i.e, different information sources contain content in different languages [20]. A common method to handle MLIR involves a query translation step in which a user submitted query is translated to the languages of the information sources using resources and tools such as multilingual dictionaries and machine translation systems [2, 5, 21]. The current challenge of MLIR is to present search results from multiple sources in an order that will be useful to the user. Two common multilingual search results presentation formats have been used – interleaving results based on an algorithm or presenting results in different languages in different tabs or pages [30]. Interleaving results requires a single ranked document list to be created, which creates the problem of result merging in Information Retrieval (IR).

Several algorithms for result merging have been proposed: (i) using normalised similarity scores [20]; (ii) downloading and translating retrieved documents [29]; and (iii) machine learning approaches

that learn to rank documents using user preferences [14]. Approaches that rely on language resources and tools are currently not appropriate for RSLs. Additionally, downloading documents to perform machine translation is a computationally expensive task. In the context of interleaving search results using normalised scores, several algorithms have been proposed [20, 25]. Oztekin *et al.* [25] used a Round Robin approach to rank results from a federated search, i.e., ranking based on the order in which they appear from each source – all the first results, followed by all the documents ranked second, followed documents ranked third, and so on. However, some studies have used raw scores, i.e., scores generated by the ranking algorithm for each source or language using methods such as cosine similarity [6, 20]. This method assumes the scores of the documents from different collections are comparable and merges the results into a single list using the original scores. However, studies using raw scores have shown mixed results: comparison of results with new algorithms obtains better or comparable results [20]. Hence, the need for more research in this area to understand which merging strategies work better for different contexts.

3 SYSTEM OVERVIEW

The BantuWeb portal is made of up three components: a Web portal to allow users to interact with the digital library, a Web crawler to get content from the Internet, and a search feature to allow users to find content from the digital library. The Web portal acts as a gateway for end-users to access and interact with the system. Users interact with the resources available on the Web portal or create their own with registered accounts. The system overview is shown in Figure 1.

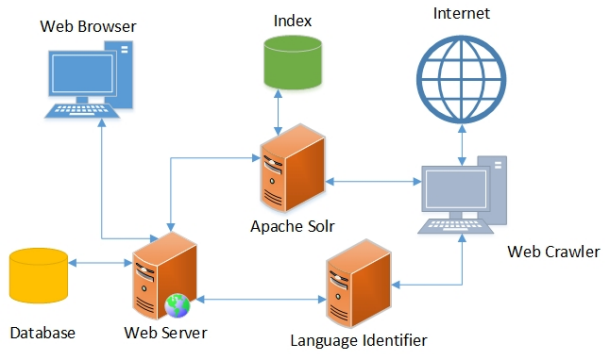

Figure 1: BantuWeb System Overview

3.1 Web Crawler

The digital library consists of Web pages crawled from the Web and content contributed by users. The language focused crawling integrated a Web crawler and a language identification tool. A language identification tool is used to identify the language of the documents before it is indexed. A Java implemented Web crawler, crawler4j[1], was used to crawl the Web for documents written in any of the nine languages. A language identification tool – National Centre for Human Language Technology (NCHLT) South African Language Identifier (SALID) [2] – was used to identify the language of the documents during Web crawling and user document upload. Apache Solr version 6.2.0 was used to index and search documents. The web crawler used a seed list of URLs extracted from searches made to Bing using 1000 words for each supported language that were created from NCHLT text corpora. Each language had its own collection and was indexed separately. Table 1 shows the number of documents crawled for each language. The Web crawler retrieved

Language	Collection Size
isiNdebele	2 127
isiXhosa	21 223
isiZulu	104 148
Sesotho	16 911
Sepedi	26 718
Setswana	16 275
siSwati	7 358
Tshivenda	3 397
Xitsonga	9 839

Table 1: Number of documents crawled from the web by language

a total of 207,996 Web pages across the 9 supported South African languages. A subset of the acquired collection was manually evaluated by 45 multilingual speakers within the multilingual search evaluation task in Section 4.5 and reported that 18 Web pages had an incorrect language out of 2,270 Web pages were viewed.

3.2 Search Platform

The search platform allows users to search for user uploaded content and content harvested from the Web. Apache Solr is used to search for the relevant Web pages indexed in its collections. The collections being searched are specified by the user's primary language and the language(s) which they have chosen to retrieve the results in. The search query that a user submits, is translated into the selected language(s). The translations are handled by the Google Translate API, using the selected primary language as the source language of the translation and the other selected language(s) as target language(s). The translated queries are submitted to Solr to query the appropriate collection of Web pages or documents. Each language collection returns its own list of results, which are merged into a single result list and presented to the user. Figure 2 shows the search interface for submitting queries.

Two algorithms for merging and ranking results were investigated with the portal: blind interleaving and smart interleaving. Blind interleaving ranks results based on their initial similarity scores in a round robin fashion. The first result from the language of the query becomes the first result to be displayed, the second result is the first result from the second language, until all the languages and results are covered. A smart interleaving algorithm merges the results from each collection using normalised scores. A

[1]https://github.com/yasserg/crawler4j

[2]https://rma.nwu.ac.za/index.php/resource-catalogue/nchlt-south-african-language-identifier.html

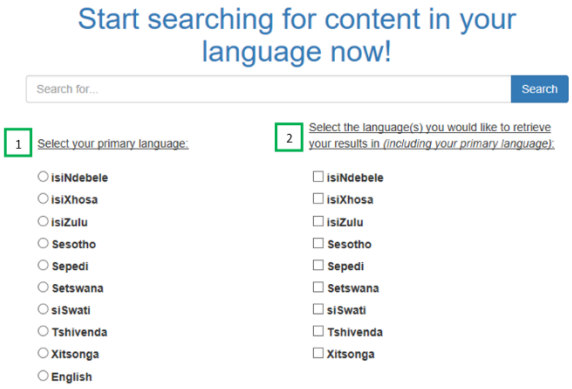

Figure 2: Search interface for submitting user queries. (1) selects the language of the query and (2) adds languages that should be searched for.

re-ranking step takes the initial top ten results from each result set and calculates the similarity scores as if the collection contains only ten documents per language. Therefore, the collection size of each result list is normalized to ten in the re-ranking step. Only the top ten results from each collection are being re-ranked for the reported experiments as we are only displaying top ten results to the user: this is due to the fact that users expect the result they are looking for to be higher up in the list and tend to not look further than the first ten results [18, 25]. The re-ranking algorithm uses Solr's default tf-idf based scoring method to calculate the normalised scores.

3.3 Web Portal Features

The Web portal has features that are accessible to the public and to registered users only. Non-registered users are able to navigate user uploaded documents as well as search for content. Users register to use the following features: upload documents and review documents. Also, registered users have gamification features. Users are rewarded with virtual points for completing tasks on the Web portal. Users get 60 points for uploading a document and 15 points for reviewing a document uploaded by other users. Users also receive 1 point for each time their document is viewed by another user.

Using virtual points as a ranking method, a leaderboard lists all registered users. The leaderboard aims to invoke competition between users on the system to motivate them to contribute content and compete for the top position on the leaderboard. Users have a level based on how many points they have earned. The levels get increasingly harder to reach as the point difference between levels increases. The level feature is used as a progression indicator as well as to allow users to work towards a goal when earning points. Users are notified and reminded of the gamification in various ways. When users complete tasks, pop-up notifications announce how many points the user has earned. A notification bar reminds the user how many uploads or reviews are needed to reach the next level. This feature aims to encourage the user to contribute by reminding them of their progression.

The Web portal or front-end was developed using Bootstrap as a responsive mobile-friendly User Interface layer. The Web portal is localized using the ng-translate AngularJS package, allowing the user interface to change to any of the nine South African RSLs by selecting the desired language. The translations for the Web portal elements are defined in JSON files that contain the lookup values for each language. The back-end was built using Python Django to create an API that can be queried for content using a RESTful architecture.

4 EXPERIMENTS AND RESULTS

Two studies were done to evaluate gamification features and usability of the portal as well as multilingual search. Convenience sampling was used to recruit participants of the study and participants were drawn from the university population. The eligibility criteria for the experiments were to be at least 18 years of age and to be a speaker of least one of the South African RSLs. 45 participants took part in the studies, and an additional 5 participated in a pilot study. Participants performed controlled experiments and were asked to complete questionnaires before and after the experiments using the LimeSurvey tool. All experiments took at least 90 minutes and participants were compensated with money based on the benchmarked time of the experiment.

4.1 Demographics

The initial task in the study was a demographics survey to capture gender, occupation, age and the languages spoken by a participant. This information is not tied to individual participants, but is used to understand the demographics of the participants when examining the results. Information about their search experience using any of the languages that they speak was captured. Search experience questions covered whether the participants search for content in South African RSLs and how satisfied they were with the results.

Participants Experiments were conducted by a group of participants composed of 13 male and 32 female students, all between the ages of 18 and 35. IsiXhosa was the most represented language with isiZulu being second. These two languages have the most representation, as they are the most spoken languages among the South African RSLs.

4.2 Searching Habits

Participants were asked about their searching habits in the South African RSLs. 58% of the participants indicated that they have searched for online content in RSLs with the majority of them searching on a monthly basis, as seen in Figure 3. Only six participants search frequently: weekly and daily. 42% of the participants had never searched for content online using the investigated languages. The reasons for not searching for content using RSLs are shown in Figure 4. 65% of participants indicated that they are not motivated to search for content online in RSLs because they cannot find content, exposing the reason why 42% of participants do not search for content online in RSLs.

However, responses from other questions indicated that participants consume content written in African languages in other media, such as books and magazines. Only 29% indicated to not consume any content in these languages in other media.

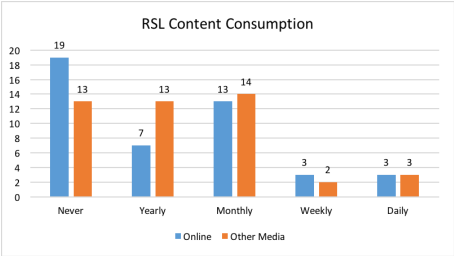

Figure 3: Content consumption in RSLs through online search engines and other media

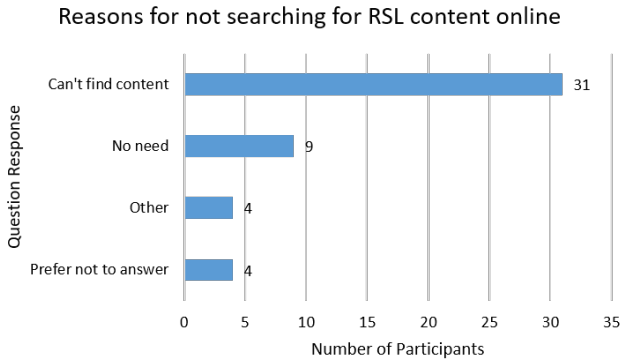

Figure 4: Reasons why participants do not consume online content in RSLs

Among the participants who indicated to have searched for content online in RSLs before, the majority only find results occasionally, as seen in Figure 5. When considering the quality of the content found online, 46% of the participants indicated that the results were poor or bad, as shown in Figure 6. From these recorded results (shown in Figure 3, 4, 5 and 6), it can be concluded that there is an interest to consume content in RSLs, yet users are not satisfied with the results and how often they find results.

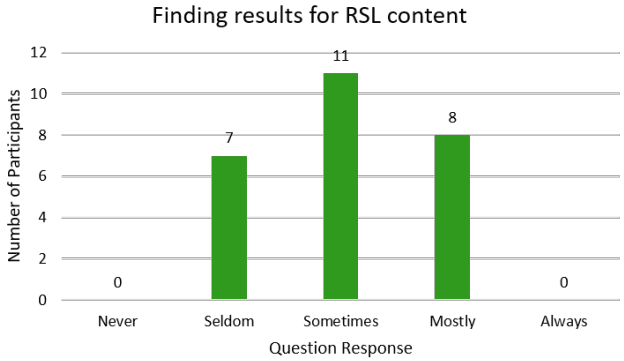

Figure 5: How often results are found for online searches in RSLs

Figure 6: The rated quality of search results for online searches in RSLs

4.3 Usability Evaluation

The Web portal was evaluated in terms of the gamification features and usability. Participants were given a task to execute and thereafter, they were asked to complete two questionnaires. The usability evaluation was conducted using the System Usability Scale (SUS) [3, 4], to gauge the usability of the BantuWeb system. The SUS identifies any usability flaws present in the system. The usability is also a factor that can hinder the gamification evaluation, by preventing users from using the system as intended. Participants answer 10 questions on a Likert scale of 1 (Strongly Disagree) to 5 (Strongly Agree). Bangor *et al.* [1] have done extensive research about the SUS, as a means of measuring usability. Bangor *et al.* [1] conducted research into the individual questions of the SUS questionnaire and also formulated a scale on which SUS scores can be interpreted. According to Bangor *et al.* [1] a score below 50 is an indication of serious usability issues, a score in the low 70's is considered good and a score above 85 is considered excellent. These categories are used to judge the score received for the BantuWeb portal.

The system scored a SUS score of 70.1 out of 100, which is considered to be an almost good score and falls within the acceptable range [1]. In Figure 7, the variation in scores is plotted, with the minimum rating being 47.5 and the maximum being 90. The mean score of 70.1 suggests that the system has a good usability rating and that there are improvements to be made. The minimum score suggests that some parts of the system were not usable, with some ratings falling in the 45 to 55 region, indicating poor usability. The standard deviation of the SUS scores was 10.4. This standard deviation is still a good result as the deviated scores still fall within acceptable SUS scores. Therefore, the usability of the system was rated well and implies no major faults in the user interface system. This means that the usability is an unlikely factor that influenced the performance of the gamification features. Apart from the SUS evaluation, a number of participants indicated that the system was easy to use and they enjoyed using it. However, participants indicated that some of the interfaces were cluttered.

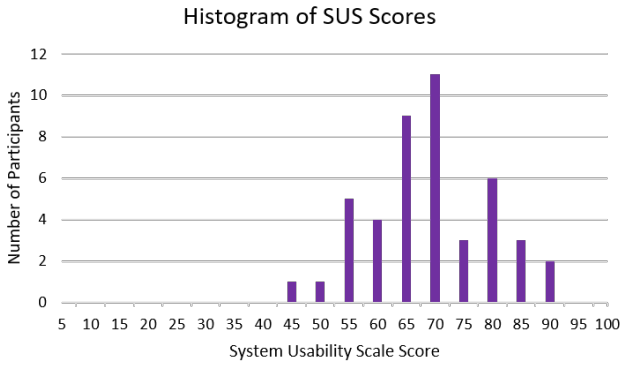

Figure 7: Variation of SUS scores recorded for usability evaluation of the user interface of the web portal

4.4 Motivation Evaluation

Following the SUS evaluation, participants were asked to answer questions to determine if the system features motivated them to contribute content. The questions were broken down into four categories of evaluation : (i) level feature; (ii) leaderboard feature; (iii) user experience on whether participants felt motivated and rewarded while contributing content; and (iv) what motivated the participants to contribute.

The evaluation of the level feature consisted of six questions. These questions asked whether the participants knew about the level feature and its progression, participants wanted to improve their level and if their level motivated them to contribute content to reach the next level. Similarly, the leaderboard feature was evaluated using six questions. This category had questions to know if a participant was aware of the leaderboard feature and if their ranking motivated them to contribute content. The third category of the evaluation investigated whether participants felt motivated and rewarded while contributing content. The questions for evaluating the leaderboard, level, and motivation and rewards categories are listed in Table 2. Responses in these categories were averaged for individual questions and for each category. The results shown in Figures 9, 10 and 11 use question codes that are in reference to Table 2.

The category evaluated what motivated the participants to contribute and included an open ended question for participants to add factors that were not included. The part had eight options which were to be ranked by the participants by how much they were motivated by the factor (from the highest to the lowest), which include: *I am being paid, To contribute more than others, Getting points, Reaching the next level, To help others using this system in the future, I think it is an important project, I want to increase the presence of my language(s)* and *I want others to learn about my language(s).* The questions were constructed to be similar to the SUS questions in grammar and wording. This part was analysed by assigning weights to the ranked results. The result ranked first is weighted with 8 points, because there are 8 options, where the last ranked element receives 1 point. The points are summed across all results to give a final list of ranked results.

Table 2: Question codes for motivation evaluation

Code	Question
GE11	I was aware of my level at all times
GE12	I was aware of how many points I still needed to reach the next level
GE13	I wanted to improve my level when I visited the Website
GE14	I felt motivated to contribute because I wanted to reach the next level
GE15	I uploaded documents to improve my level
GE16	I reviewed documents to improve my level
GE21	I was aware of my position on the leaderboard when using the Website
GE22	I checked my position on the leaderboard at least 3 times
GE23	I wanted to rank higher than other users featuring on the leaderboard
GE24	I felt motivated to contribute because of my leaderboard position
GE31	I uploaded documents to improve my leaderboard position
GE32	I reviewed documents to improve my leaderboard position
GE33	I felt motivated to upload documents
GE34	I felt motivated to review documents
GE35	I felt rewarded after uploading documents
GE36	I felt rewarded after reviewing documents

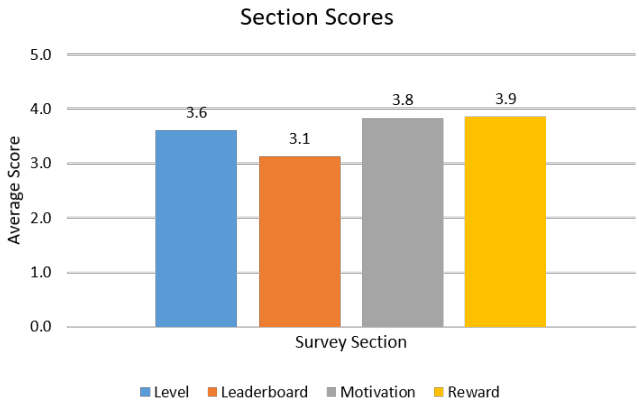

Figure 8: Results for the four categories that were evaluated in the context of motivation, namely, level, leaderboard, user experience on being motivated and rewarded; and (iv) ranking of motivation factors

In reference to Figure 8, the leaderboard feature performed the worst with an average score of 3.1/5 across all questions. The motivation and reward parts performed the best with high scores showing that users were motivated to contribute content and that they felt rewarded for contributing. The level feature performed better than the leaderboard feature with an average of 3.6/5. The

category scores indicate that the level gamification feature received a higher positive response than the leaderboard feature.

To more closely evaluate how gamification features rank in terms of motivating users to contribute content in RSLs, a Wilcoxon-Signed-Rank Test was done on the responses to GE14 and GE24, as these questions directly evaluated how the features motivated users. The medians of GE14 and GE24 were 3.7 and 3.3, respectively. A Wilcoxon Signed-rank test shows that there is a significant difference between the responses of the two questions ($W = 243$, $Z = 2.8868$, $p < 0.05$, $r = 0.032$). This test shows that there is a significant difference between the scores recorded for GE14 and GE24, meaning the level feature performed better than the leaderboard feature. Questions GE13 - GE16 received high scores (> 3.5), indicating that participants wanted to improve their levels and that levels motivated them to contribute content. Scores above 3.5 indicate that they are aligned with the 'Agree' Likert scale answer, concluding that the participants did agree that the level feature motivated them. This result means that the level feature was successful in motivating users to contribute content on the Web portal. However, reviewing other users contributed content to improve level was rated higher than uploading own content to improve one's level. This is likely due to the ease of reviewing documents compared to creating and uploading documents.

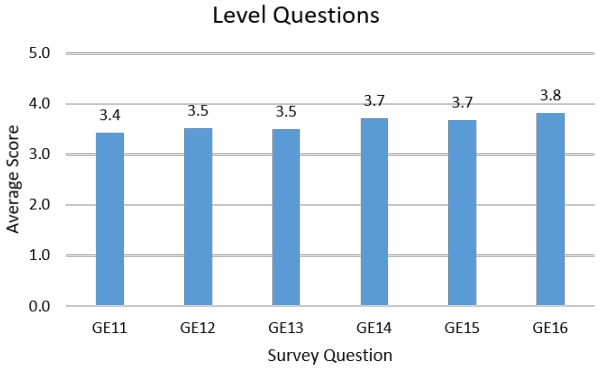

Figure 9: Results of responses to questions evaluating the level feature

The individual scores for the leaderboard feature in Figure 10 received less positive results. A clear result is that participants did not check their ranking on the leaderboard during the experiment, as this question received an average response of 2.5. The constrained time frame could be a limiting factor for the leaderboard feature because the users are too busy with the tasks to fully explore and use the leaderboard feature.

The question addressing the competition between users received an average response of 3, meaning Neutral in the evaluation. This result indicates that participants did not purposefully compete with other participants, but rather did not experience a sense of competition as they did not agree or disagree with the statement. Participants were slightly motivated by their leaderboard position as seen by results to question GE24, GE31 and GE32, as these had value >3, yet are aligned with the Likert scale of 'Neutral'. The response of motivation being 'Neutral' again indicates that participants were not influenced by the leaderboard feature as expected.

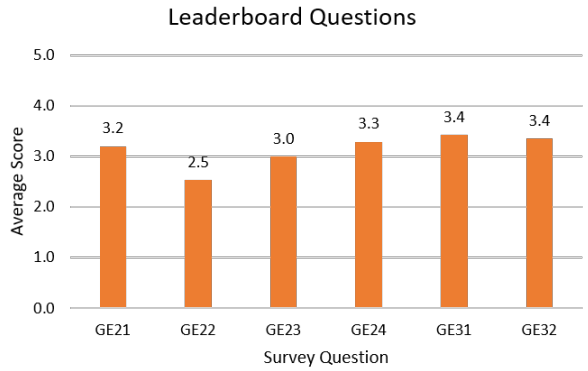

Figure 10: Results of responses to questions evaluating the leaderboard feature

The motivation and reward sections got positive feedback (>3.5), as seen in Figure 11. Participants felt less motivated to contribute through reviews and felt less rewarded by doing reviews. These lower results are presumably because the review feature is used less in the tasks. The difference could also be explained because creating a document is a bigger contribution than reviewing a document both in the effort required and virtual reward received. The increased effort and reward could motivate users more as the extra effort is more rewarding than a lot of little contributions through reviews.

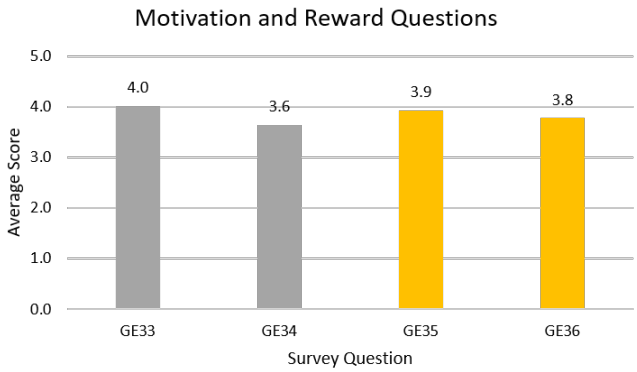

Figure 11: Results of responses to motivation and reward category

The last category of ranking motivation factors was divided into three factors: gamification motivation, monetary motivation and community motivation. The gamification motivation included the level, virtual points and leaderboard motivational factors. The second motivator category was receiving a monetary reward for the

experiment. The community category is composed of the desire to increase the presence of a RSL, to help future users of the system, to educate other users in a RSL and to help the project become a success because it is deemed important. This last factor was ranked higher by participants, i.e, participants were motivated to contribute to increase the presence of languages. An interesting result is that all four community motivators were ranked the highest, followed by the monetary reward and then the gamification features. The community motivators ranking above all is unexpected, especially because Havenga *et al.* [17] found that competition was valued more than community and collaboration. Havenga *et al.*'s [17] study was conducted five years ago, which might be a reason for different results. Recent events in South Africa have led to a strong call for decolonization across all sectors and could be a reason why community values are regarded as more important than competition. The majority of content written by the participants was about decolonization and the student protests that were happening during that time across the country. Another reason the results may differ is that the format of contributions between the two studies differ. Havenga *et al.* [17] worked with photographs, where this project works with text. Text contributions could be more motivational because of community orientated goals than photograph contributions are, yet this would need to be evaluated further.

Another interesting observation is that the categories are grouped closely together, where the community contributors all ranked fairly close together and relatively far above the monetary reward and even further from the gamification features. The same pattern can be seen in the gamification motivators, where all three have similar point scores.

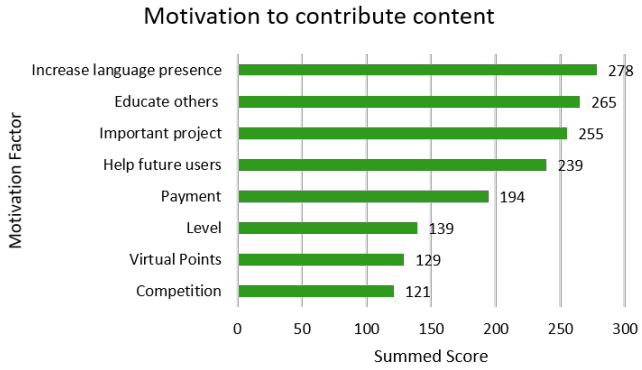

Figure 12: Motivational factors for content contribution

The three gamification motivators in Figure 12 rank below the monetary reward for participants, which suggests that the money was more motivating than the gamification features. Of the three components, the level was rated the best contributer, suggesting that the level gamification feature motivated users the most to contribute content among the gamification features. Virtual points ranked slightly higher than competition, indicating that the idea of virtual rewards was more motivating than the competition between users. The competition being less effective than the virtual rewards is the opposite of what Havenga *et al.* [17] found. The factor of payment was something that needed to be considered as a bias in the experiments, yet only ranked as fifth. This result is good, because it shows that participants did contribute for reasons other than the payment. A small number of participants did indicate verbally and through the comment section that they would have contributed without payment. In addition, the majority of participants verbally engaged with the researchers after the experiment to find out how to access the portal outside of the experiment and when it would be accessible. The influence of the monetary reward, however, can not be accurately measured as there is no control group to compare the results to.

4.5 Multilingual Search Evaluation

Participants were asked to formulate a search query on some topics covered in the collection. These queries were used to evaluate the quality of search results using the two methods for merging and ranking search results. The task was divided into three phases: query formulation, monolingual retrieval and multilingual retrieval using isiXhosa, isiZulu and Sesotho and, lastly, answering a questionnaire. Only three languages were used in the retrieval task due to the coverage of languages in Google translate as well as the size of the corpus that was crawled for the individual languages. In each retrieval task, participants marked the relevancy score of each result, i.e., the top ten results the system returned for the query, and the system captured these judgments. Participants were also asked to answer a questionnaire about their search experience on the portal. To compare the quality of results returned by the two interleaving methods, two measures of overall relevancy evaluation were used – namely, Mean Average Precision (MAP) and Normalized Discounted Cumulative Gain (NDCG). The Mean Average Precision measure uses binary relevance judgments for each result, i.e., a participant indicates whether a Web page or document is relevant or irrelevant. These relevance evaluation measures make the assumption that the relevance of each Web page or document is independent of the other Web pages or documents that occur in the result list [15].

$$P@n = \frac{\#relevant\ Webpages/docs\ in\ top\ n\ results}{n} \tag{1}$$

The Average Precision (AP) of the results for a query is calculated by getting the average of the P@n for all values of n [15].

$$AP = \frac{\sum_{n=1}^{10} (P@n)}{n} \tag{2}$$

MAP is therefore calculated by taking the arithmetic mean of the AP values for each query.

The Normalized Discounted Cumulative Gain measure was developed by Järvelin and Kekäläinen, which allows for graded relevance judgements as opposed to the binary relevance judgements used in MAP [18]. Based on the fact that users are less likely to look at the results further down the list - a rank-based discount factor is removed from the score of the Web pages or documents found further down the result list [18]. The average DCG of all the queries

over all sources were calculated.

$$DCG = \sum_{i=1}^{10} \frac{V[i]}{\log(i+1)} \tag{3}$$

A single NDCG value, α-NDCG was calculated for all the queries submitted by the participants. This is calculated by taking the arithmetic mean of the NDCG values for each query [15].

$$\text{NDCG} = \frac{\text{DCG}}{\text{IDCG}} \tag{4}$$

Where IDCG (Ideal Discounted Cumulative Gain) is the ideal ordering of the DCG vector.

The performance measures, MAP, DCG and NDCG, were used to measure the relevancy of the top ten results returned by the two interleaving methods. In Table 3 are the MAP, DCG and α-NDCG values of the query results judged by the participants. The expectation was that smart interleaving will produce result lists with higher relevancy scores to the user than blind interleaving. As blind interleaving is the simplest interleaving method, it is used as a base line for the experiment, and improvements are expected from the smart interleaving method. The hypothesis is that smart interleaving will produce result lists with higher relevancy scores to the user than blind interleaving. The results indicate that there

	Blind Interleaving	**Smart Interleaving**
MAP	0.3047	0.3412
DCG	2.0798	2.3705
α-NDCG	0.5961	0.6224

Table 3: MAP, DCG and α-NDCG values for blind and smart interleaving

may be a difference in performance for the queries. The results were calculated from 45 multilingual search queries, with each query being repeated for both blind and smart interleaving. As seen in Table 3, there is an increase in the MAP, DCG and α-NDCG scores between the blind and smart interleaving methods. The increase in the MAP scores are not statistically significant ($p > 0.05$) and therefore does not reject the null hypothesis. On the other hand, DCG scores, which uses graded relevancy judgments, are statistically significant ($p < 0.05$) and therefore reject the null hypothesis.The DCG scores show that the overall relevancy of the results returned in the final result list is higher for smart interleaving than blind interleaving.

The search feature does not take into account that the Bantu languages are agglutinative languages [28]. Therefore, the grammatical information attached to the word, by prefixes and suffixes, were not removed when the documents were indexed and the queries were submitted. This might have hindered the documents or Web pages that would have matched with the query.

5 DISCUSSION

The addition of gamification features, such as a leaderboard, virtual points and levels, as a motivational strategy, provided positive results. Additional motivational variables were uncovered in the evaluation that are a possible area of study for future work. The gamification features motivated users to contribute in the environment presented by this study. However, there was no control group and the study was done in a short period of time (2 hours). This is a limitation of the study, as the results cannot be compared to a system without the gamification features. A mean score of 3.5 is calculated from the questions that measured the motivation inspired by the gamification features. This mean score aligns with the 'Agree' Likert scale answer, which suggests that the gamification features did motivate users to contribute content. Therefore, the results indicate that the gamification features had a positive effect, bearing in mind the limitation of not having a control group. The level gamification feature proved to be the more effective gamification feature in the evaluation. The level feature ranked higher in the two evaluations that were made on gamification features. Apart from these two results, a Wilcoxon test was done and showed a significant difference ($p < 0.05$) between the two features. The overwhelming positive feedback from participants and the search habit results show the desperate need for a system such as this and how much RSL speaking South African's would use a system like this.

Two merging techniques were compared: blind and smart interleaving. Using MAP, a binary relevance measure, the results do not show any statistical significance in the improvement in smart interleaving over blind interleaving. Similar results are obtained for NDCG. Using DCG, a graded relevance measure, the results do show a statistically significant improvement in the relevance of results returned by smart interleaving over blind interleaving. DCG is affected by variations in the number of documents available for a query, i.e., queries with more relevant results have a high DCG. NDCG normalises this effect by dividing the DCG with IDCG, i.e. the DCG of the best possible ranking. NDCG and MAP measure the overall goodness of the ranking and the results obtained are similar. This is not surprising as normalised weighting scores have shown in other studies to not improve quality of retrieval [29]. However, DCG shows that there may be some variations in the number of relevant results and the two ranking systems were unable to balance this aspect.

6 CONCLUSION

Overall, the gamification features had a positive impact on the BantuWeb portal and motivated users to contribute in their RSLs. It has been learned that in a Web based digital library, the level gamification feature motivates users to contribute the most. Further, it seems that the motivation behind content creation in the given context is influenced by community orientated goals rather than competition. Future work includes studying the effects of gamification on content creation over a longer duration of time with more participants, which has not been possible within the scope of this study. A bigger user base would reveal more patterns in content creation. A last aspect that can be implemented further is to allow more media to be added to the Web portal, such as videos, pictures and audio. Although these would be harder to index for searching purposes, they would add valuable content towards the preservation of the South African RSLs.

The two ranking algorithms, blind and smart interleaving, had no significant difference in performance, i.e., ranking highly relevant

documents to the query. Future work in this area could include morphological analysis techniques such as stemmer or n-grams, spell checking on the user's query, and a machine readable list of synonyms for each language to able able to expand queries.

Finally, BantuWeb is planned to be used for further research in IR for South African Languages. Data generated through search by users on the portal can be used to train models for ranking algorithms. Interestingly, the library can be used as a live laboratory to test new tools and algorithms with real users for a longer period of time.

ACKNOWLEDGMENTS

This research was partially funded by the National Research Foundation of South Africa (Grant numbers: 85470 and 88209) and University of Cape Town. The authors acknowledge that opinions, findings and conclusions or recommendations expressed in this publication are that of the authors, and that the NRF accepts no liability whatsoever in this regard.

REFERENCES

[1] Aaron Bangor, Philip T. Kortum, and James T. Miller. 2008. An empirical evaluation of the system usability scale. *International Journal of Human–Computer Interaction* 24, 6 (2008), 574–594.

[2] Shane Bergsma, Paul McNamee, Mossaab Bagdouri, Clayton Fink, and Theresa Wilson. 2012. Language Identification for Creating Language-specific Twitter Collections. In *Proceedings of the Second Workshop on Language in Social Media (LSM '12)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 65–74. http://dl.acm.org/citation.cfm?id=2390374.2390382

[3] John Brooke. 1996. SUS: A "quick and dirty" usability scale. In *In Usability evaluation in industry*, P. W. Jordan, B. A. Weerdmeester Thomas, and I. L. McClelland (Eds.). London: Taylor & Francis, 189–194.

[4] John Brooke. 2013. SUS: A Retrospective. *Journal of Usability Studies* 8, 2 (feb 2013), 29–40. http://dl.acm.org/citation.cfm?id=2817912.2817913

[5] Carlos Castillo. 2005. Effective web crawling. In *ACM SIGIR Forum*, Vol. 39. ACM, 55–56.

[6] Catherine Chavula and Hussein Suleman. 2016. Assessing the Impact of Vocabulary Similarity on Multilingual Information Retrieval for Bantu Languages. In *Proceedings of the 8th Annual Meeting of the Forum on Information Retrieval Evaluation (FIRE '16)*. ACM, New York, NY, USA, 16–23. DOI: http://dx.doi.org/10.1145/3015157.3015160

[7] Peng Chu and Anita Komlodi. 2017. TranSearch: A Multilingual Search User Interface Accommodating User Interaction and Preference. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*. ACM, New York, NY, USA, 2466–2472. DOI: http://dx.doi.org/10.1145/3027063.3053262

[8] Gilles-Maurice De Schryver and Daan J Prinsloo. 2000. The compilation of electronic corpora, with special reference to the African languages. *Southern African linguistics and applied language studies* 18, 1-4 (2000), 89–106.

[9] Paul Denny. 2013. The Effect of Virtual Achievements on Student Engagement. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 763–772. DOI: http://dx.doi.org/10.1145/2470654.2470763

[10] Sebastian Deterding, Miguel Sicart, Lennart Nacke, Kenton O'Hara, and Dan Dixon. 2011. Gamification. Using Game-design Elements in Non-gaming Contexts. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems (CHI EA '11)*. ACM, New York, NY, USA, 2425–2428. DOI: http://dx.doi.org/10.1145/1979742.1979575

[11] Marcel Diki-Kidiri. 2008. *Securing a place for a language in cyberspace*. Information Society Division Communication and Information Sector UNESCO Paris, France.

[12] Rosta Farzan, Joan M DiMicco, David R Millen, Beth Brownholtz, Werner Geyer, and Casey Dugan. 2008. When the experiment is over: Deploying an incentive system to all the users. In *symposium on persuasive technology*.

[13] Rosta Farzan, Joan M. DiMicco, David R. Millen, Casey Dugan, Werner Geyer, and Elizabeth A. Brownholtz. 2008. Results from Deploying a Participation Incentive Mechanism Within the Enterprise. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. ACM, New York, NY, USA, 563–572. DOI: http://dx.doi.org/10.1145/1357054.1357145

[14] Wei Gao, Cheng Niu, Ming Zhou, and Kam-Fai Wong. 2009. Joint Ranking for Multilingual Web Search. In *Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval (ECIR '09)*. Springer-Verlag, 114–125. DOI: http://dx.doi.org/10.1007/978-3-642-00958-7_13

[15] Rayid Ghani, Rosie Jones, and Dunja Mladenić. 2001. Mining the web to create minority language corpora. In *Proceedings of the tenth international conference on Information and knowledge management*. ACM, 279–286.

[16] Juho Hamari, Jonna Koivisto, and Harri Sarsa. 2014. Does Gamification Work? – A Literature Review of Empirical Studies on Gamification. In *Proceedings of the 2014 47th Hawaii International Conference on System Sciences (HICSS '14)*. IEEE, 3025–3034. DOI: http://dx.doi.org/10.1109/HICSS.2014.377

[17] Michelle Havenga, Kyle Williams, and Hussein Suleman. 2012. Motivating Users to Build Heritage Collections Using Games on Social Networks. In *The Outreach of Digital Libraries: A Globalized Resource Network: 14th International Conference on Asia-Pacific Digital Libraries, ICADL 2012, Taipei, Taiwan, November 12-15, 2012, Proceedings*, Hsin-Hsi Chen and Gobinda Chowdhury (Eds.). Springer Berlin Heidelberg, 279–288. DOI: http://dx.doi.org/10.1007/978-3-642-34752-8_34

[18] Kalervo Järvelin and Jaana Kekäläinen. 2000. IR Evaluation Methods for Retrieving Highly Relevant Documents. In *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '00)*. ACM, New York, NY, USA, 41–48. DOI: http://dx.doi.org/10.1145/345508.345545

[19] Shigeaki Kodama. 2008. Languages on the Asian and African Domains. In *Proceedings of the International Symposium on CDG*. 77–82.

[20] Wen-Cheng Lin and Hsin-Hsi Chen. 2003. *Merging Mechanisms in Multilingual Information Retrieval*. Springer Berlin Heidelberg, 175–186. DOI: http://dx.doi.org/10.1007/978-3-540-45237-9_14

[21] Marco Lui and Timothy Baldwin. 2012. langid. py: An off-the-shelf language identification tool. In *Proceedings of the ACL 2012 system demonstrations*. Association for Computational Linguistics, 25–30.

[22] Markus Montola, Timo Nummenmaa, Andrés Lucero, Marion Boberg, and Hannu Korhonen. 2009. Applying Game Achievement Systems to Enhance User Experience in a Photo Sharing Service. In *Proceedings of the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era (MindTrek '09)*. ACM, 94–97. DOI: http://dx.doi.org/10.1145/1621841.1621859

[23] Jan Alewyn Nel, Velichko H Valchev, Sebastiaan Rothmann, Fons JR Vijver, Deon Meiring, and Gideon P Bruin. 2012. Exploring the personality structure in the 11 languages of South Africa. *Journal of Personality* 80, 4 (2012), 915–948.

[24] Don Osborn. 2010. *African Languages in a Digital Age: Challenges and opportunities for indigenous language computing*. IDRC.

[25] B. Uygar Oztekin, George Karypis, and Vipin Kumar. 2002. Expert Agreement and Content Based Reranking in a Meta Search Environment Using Mearf. In *Proceedings of the 11th International Conference on World Wide Web (WWW '02)*. ACM, New York, NY, USA, 333–344. DOI: http://dx.doi.org/10.1145/511446.511490

[26] Sean Packham and Hussein Suleman. 2015. Crowdsourcing a Text Corpus is not a Game. In *Digital Libraries: Providing Quality Information: 17th International Conference on Asia-Pacific Digital Libraries, ICADL 2015, Seoul, Korea, December 9-12, 2015. Proceedings*, Robert B. Allen, Jane Hunter, and Marcia L. Zeng (Eds.). Springer International Publishing, Cham, 225–234. DOI: http://dx.doi.org/10.1007/978-3-319-27974-9_23

[27] Carol Peters, Martin Braschler, and Paul D. Clough. 2012. *Multilingual Information Retrieval - From Research To Practice*. Springer Berlin Heidelberg. DOI: http://dx.doi.org/10.1007/978-3-642-23008-0

[28] Hans Jiirgen Sasse. 1992. Theory of language death. *Language death: Factual and theoretical explorations with special reference to East Africa* 64 (1992), 7.

[29] Luo Si, Jamie Callan, Suleyman Cetintas, and Hao Yuan. 2008. An Effective and Efficient Results Merging Strategy for Multilingual Information Retrieval in Federated Search Environments. *Information Retrieval* 11, 1 (Feb. 2008), 1–24. DOI: http://dx.doi.org/10.1007/s10791-007-9036-6

[30] Ben Steichen and Luanne Freund. 2015. Supporting the Modern Polyglot: A Comparison of Multilingual Search Interfaces. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 3483–3492. DOI: http://dx.doi.org/10.1145/2702123.2702541