Knowledge-Based Systems 112 (2016) 127–141

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Use of textual and conceptual profiles for personalized retrieval of political documents

Eduardo Vicente-López, Luis M. de Campos*, Juan M. Fernández-Luna, Juan F. Huete

Departamento de Ciencias de la Computación e Inteligencia Artificial, E.T.S.I.I.T., CITIC-UGR, Universidad de Granada, 18071-Granada, Spain

ARTICLE INFO

Article history:
Received 24 May 2016
Revised 4 August 2016
Accepted 7 September 2016
Available online 8 September 2016

Keywords:
User profile
Personalization
Information retrieval
Privacy-enhanced
E-Government

ABSTRACT

The amount of information we are exposed to on a daily basis is increasing exponentially. Besides, Information Retrieval Systems (IRSs) return the same results for a given query regardless of who submitted it. In order to address the problems of finding useful, relevant information and adapting the results to the user, the use of personalization techniques is now more necessary than ever. They are not, however, particularly popular in live environments as users remain unconvinced about their reliability and, more importantly, are concerned about privacy issues. We have developed and compared six generic user profile representations in order to improve the personalization process and address the problem of privacy. We propose a new weighting scheme to build the profiles and a new personalization technique to join the advantages of some of the previous profiles. A comprehensive evaluation study of the proposed generic user profiles was performed and this revealed very good personalization performance results and some interesting conclusions about their use in a political context, more specifically with official documents from the Andalusian Parliament.

1. Introduction

Nowadays, most information is created and exchanged in digital format with an exponential increase in recent years [9]. The e-Government framework is a specific yet important area of application. In this article, we focus on the parliamentary context in which vast amounts of information have been published. An abundance of information is pointless, however, if citizens and politicians are unable to find the relevant documents that match their information needs, generally related to problems affecting their daily lives.

Most parliaments have two main official publications: the records of the parliamentary proceedings (plenary and committee sessions) and the official bulletins. In a plenary session, political groups will present their proposals which are then debated and put to the vote. Committee sessions, on the other hand, cover particular fields such as agriculture, education or the economy. Each parliamentary proceedings document contains the full transcriptions of the speeches given by the members of the parliament in each session. The main component of these documents is the *initiative*, which presents a detailed discussion of a specific issue. Each initiative is then manually tagged by expert parliamentary documentalists with one or more subjects from the EUROVOC[1] thesaurus in order to classify the content.

Our research group has collaborated with the Andalusian Parliament since 2005 and has had access to their official publications. These are in XML format and some comprise the document collection used in the evaluation process in this article. As a result of this collaboration, we have built the *Seda*[2] IRS [11] in order to improve public access to these official parliamentary documents.

In most cases, traditional IRSs are used to access to parliamentary documents facing the following main problems: a large amount of information is available, users tend to formulate short and ambiguous queries [30], and little is known either about the users or their information needs except for their query keywords. As a result, IRSs tend to retrieve the same results for a given query regardless of the user. This issue is known as the *one size fits all* problem and personalization [4,5,18,31,36] offers a possible solution. In personalization, both the user and the query are important in the retrieval process. The main objective of personalization is to retrieve results which best suit the user to better satisfy the user's specific information needs, thereby improving the user satisfaction with the IRS.

* Corresponding author. Fax: +34 958243317.
E-mail addresses: evicente@decsai.ugr.es (E. Vicente-López), lci@decsai.ugr.es (L.M. de Campos), jmfluna@decsai.ugr.es (J.M. Fernández-Luna), jhg@decsai.ugr.es (J.F. Huete).

[1] http://eurovoc.europa.eu/
[2] http://irutai2.ugr.es/SEDA/

http://dx.doi.org/10.1016/j.knosys.2016.09.005

Although we did intend to introduce various personalization features into *Seda*, we encountered certain privacy issues since Parliament does not allow any personal data to be collected about members of parliament or the public. This is the norm rather than the exception and is increasingly becoming a major barrier to personalized IRSs [28]. According to [19], approximately 85% of users are concerned about the privacy or security of their online personal information, 90% have at some time refused to provide online personal information and 35% supply false online personal information.

This article therefore has two main objectives: 1) to provide an alternative option for implementing personalization in privacy-constrained environments, with a good performance and without the need to collect any personal information. This would be achieved by means of *generic profiles* which are learned from the document collection content; in our particular case from the Andalusian Parliament *committee sessions* which cover specific areas of interests, and 2) to be able to select the best representation of the previous generic profiles and configuration parameters to be used for any given personalization technique and retrieval scenario.

While these generic profiles might be considered *'unrealistic'*, since they do not represent real users, they are a valid approach [10,29] for possible users interested in certain areas. This is particularly true in our political context, where one politician may serve on several committees. Without the use of generic profiles, the user might also wish to include additional query terms to try to describe the committee session content. This might be difficult for the user and may also trigger the *query-drift* problem [43]: the inclusion of possibly unrelated terms in the original query might result in the retrieval of unexpected results which might not contain the original query terms. The user might also choose to filter out all documents that do not belong to the committee sessions but in doing so, certain relevant results might be omitted (about 25% according to our studies). Furthermore, a filtering approach is not a valid solution since relevant documents might be found not only in committee sessions but also in plenary sessions and official bulletins, and since these are not implicitly classified, they will not therefore be retrieved. As a result, the final percentage of possible relevant missed documents will be much higher than the previous value. The best approach is therefore to use some kind of personalization.

In order to achieve this article main objectives, we propose six different ways to represent the generic user profiles learned from the document collection content. These are based on general topic areas which are subsequently selected by the users according to their interests and preferences. It should be noted that these areas are quite well defined and characterize parliamentary activities. These profiles are ideal for introducing personalization into privacy-constrained environments where users are reluctant to reveal personal information, as occurs in the case of the Andalusian Parliament.

Although these profiles have been used for a political context, they can also be applied in other privacy-constrained retrieval environments. The only requirement for building our generic user profiles is to have a collection where at least a subset of its documents can be classified into different areas of interest or categories, that future users might find interesting. If this were not the case, a clustering process could be used to find clusters of similar documents according to their content, and subsequently a classification process can assign new documents to the corresponding clusters. In our case, since each document in the document collection belongs to one committee session, we have an implicitly classified document collection.

We next expose our contributions to achieve the objectives of this article: 1) the development of user profiles based only on terms from documents belonging to a given area of interest (committee sessions) irrespective of where they appear in the document; 2) the proposal of a new weighting scheme for profile items called *diffFreq* and we have shown how this is superior to the common *tf***idf* approach [26], at least in these category-based generic profiles; 3) the construction of user profiles based on the EUROVOC thesaurus subjects which are manually assigned to each initiative. Although this approach might seem promising, it has not been confirmed by our results. Since we still believed that subjects should add value, we designed different ways to obtain the maximum benefit from subjects and terms simultaneously; 4) the development of a new personalization technique which uses subjects and terms from the previous profiles with reasonably good results; 5) a *hybrid* user profile (with four variations) comprising both subjects and terms. We shall explain how each approach should be used and the results obtained; and 6) a comprehensive evaluation and comparative study of all of the proposed user profile representations.

Although generic profiles are frequently used in personalized contextual evaluation environments, e.g [29,33]., most personalized IRSs are not validated with real world experiments [42], since they are extremely difficult due to their complexity and the potential costs involved. However, these experiments are necessary to demonstrate the true effectiveness and improvements of any personalized IRS over other systems. Various efforts have been made to solve this problem, such as for example [40] where the authors present an easy automatic methodology to evaluate these personalized IRSs. With the previous comprehensive evaluation (sixth contribution) we provide the best generic profile representation and configuration parameters to be used for a given personalization technique and retrieval scenario. We have also obtained very good personalized results with a retrieval performance improvement of up to 80.17% on the non-personalized search, together with some interesting conclusions about the merits of using these generic profiles.

The remainder of the article is organized as follows: Section 2 reviews the different user profile approaches in the literature; Section 3 describes profile construction, use and the results obtained for the newly developed term-based and subject-based profiles; Section 4 explains how subjects and terms can be combined to work together, firstly with the newly developed personalization technique, and secondly with the *hybrid* profiles; Section 5 compares all of the developed profiles and presents some interesting conclusions; and finally Section 6 outlines the general conclusions of the article and proposals for future research.

2. Related work

There are three main stages to any IR personalization process: the first is to acquire and represent the user context in the user profile; the second is to exploit the user profile information in the retrieval process as well as possible; and the third is to evaluate the entire personalization process. Some additional issues may also be considered such as privacy when collecting or managing personal data [19], or different ways of presenting the personalized results [2] as simply and as intuitively as possible.

The quality of the personalized results is highly dependent on the quality of the user profile and how well its information is exploited in the retrieval process. The user profile building process is therefore an extremely important step in order to obtain good personalized results. We can see the importance of building accurate user profiles even applied to other domains such as social media [21] or IR related fields such as recommender systems [3].

The authors in [14] outline the following three main stages within the IR user profile building process related to the user information: 1) to collect it; 2) to represent it; and 3) to keep it up-

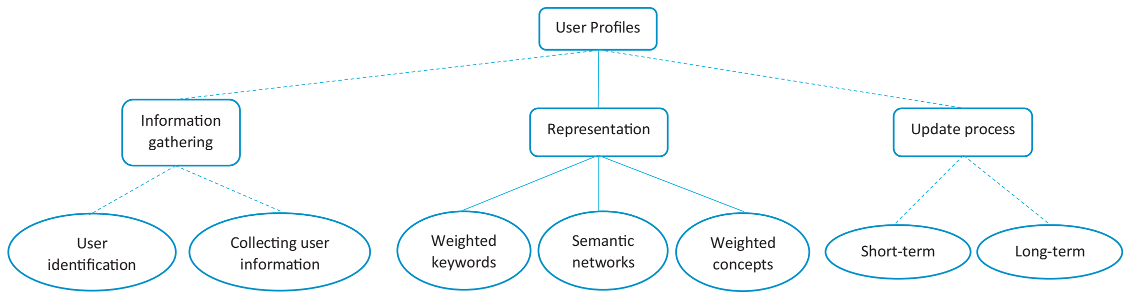

Fig. 1. Main stages of the user profile building process.

dated. With our generic user profiles we highly simplify the first and third steps. The diagram in Fig. 1 visualizes these stages.

We shall now briefly explain the first and third stages and go deeper with the second stage, since the study of different kinds of user profile representations is one of the key aspects of this article.

- *Information gathering.* The first step when building an individual user profile is to gather all the possible or necessary information about the user. In order to do so, users must be uniquely identified by the system. The three main ways of user identification are cookies, logins and software agents. The final method is not often used since users must install a program on their devices and are aware that they will be closely monitored. The best compromise is to use logins but with the additional possible use of cookies for those who do not want to register and login each time they use the system.

The information gathering process can either be performed on the user's device or on the server, and the information available will vary according to where the process is performed. This information may be collected explicitly and entered by the user or implicitly in a variety of different ways. An example of an implicit approach is [27], where the authors use implicit feedback information to perform query expansion based on previous queries, and instant result reranking based on click-through data. Generally speaking, implicit data collection places no burden on the user, and since the performance of these systems is similar to or even better than explicit systems according to [35], their use is preferred.

- *User profile representation.* In the previous step, we explained how to gather the necessary user information for building the user profile. Once we have this data, we need to define a way to represent it. This information representation will be stored in the *user profile*. This user profile will be used by personalization techniques to retrieve results that best match user interests and preferences. According to [14], there are three main representations for user profiles: weighted keywords, semantic networks and weighted concepts.

Weighted keywords. This is the most common user profile representation, the simplest to build and one of the first approaches. They require a large amount of user feedback in order to learn all the terms which represent user interest to match these interest-terms with documents to be retrieved. The keywords and their associated weights may be automatically learned from the user visited documents or directly given by the user. The keyword weights show the importance of each keyword within the profile. The main problems of this type of user profile is keyword synonymy (different words with the same or similar meanings), which may result in a recall decrease, and keyword polysemy (same word with different meanings), which may result in a precision decrease. These problems may make this kind of user profile somewhat ambiguous.

Examples of this user profile approach are Amalthaea [24], where the author learns user profiles from the web pages visited by users based on the well-known *tf*idf* approach [26]. He also allows users to explicitly provide their profiles weighting them higher than the automatically learned. In WebMate [7], the authors build user profiles comprising a vector of keywords for each user area of interest also following the *tf*idf* approach. In the case of [32], the authors propose a fine-grained search by capturing changes in user preferences. For that purpose, they build and test three different user profiles based on relevance feedback and implicit information, user browsing history, and a modified collaborative filtering. Another example is [1], where the authors extract the profile terms and weights from rss news feeds also following the *tf*idf* weighting scheme. They show and allow the user to edit this profile. Although users prefer this transparency and control they demonstrate this ability harms the system personalization performance.

Semantic networks. Within this kind of user profile representation each node represents a concept. The semantic network helps to avoid the aforementioned weighted keyword representation synonymy and polysemy problems, but they must learn the terminology (terms) associated to each concept. One example of this approach is [15], an online digital library filtering system, where the authors initially have a semantic network of unlinked concept nodes. Each concept is represented with a single, representative term for that concept. In the user profile learning process more weighted terms are associated and linked to the corresponding concepts, creating links also between concepts. Another example is detailed in [23], where a filtering interface is created to personalize the results from the Altavista search engine. The user profiles comprise three components: a header, which includes the user's personal data, a set of stereotypes and the interests for each stereotype. Another semantic network example is to be found in [29], which presents a personalized search system with ontology-based user profiles. These user profiles are built by assigning scores to user interests which have been implicitly derived from concepts of the Open Directory Project (ODP) ontology[3]. Finally, in [34] the authors propose a personalized ontology model for knowledge representation and reasoning over user profiles. This personalization model learns ontology-based user profiles from both a public knowledge base and the user's local information source.

Weighted concepts. These are similar to semantic networks since they also have conceptual nodes and relations between them, but in this case, the nodes are represented by abstract topics of interest to the user instead of terms. In contrast to semantic networks, weighted concept user profiles are trained on examples for each concept a priori, having already mapped the vocabulary and concepts. These user profiles are therefore robust to variations in terminology and are learned with much less user feedback. Meanwhile, they are also similar to weighted keyword user profiles, since they are usually represented as vectors of weighted concepts. In recent years, it is common to use a hierarchical representation of concepts which are usually derived from a taxonomy, thesaurus

[3] http://www.dmoz.org/

Table 1
Examples of the *tProf* and *sProf* profiles, for the 'agriculture and livestock' area of interest (unstemmed, with fewer decimals and translated into English) following the two different weighting schemes and using the *ss* and *ls* stopword lists.

tProf *(tf*idf - ss)*	*t* = { 0.013*agriculture 0.013*sector 0.009*agrarian 0.009*fishing 0.009*production 0.008*sir 0.008*aid 0.008*andalusia 0.007*farmer 0.007*product ... }
tProf *(tf*idf - ls)*	*t* = { 0.014*agriculture 0.014*sector 0.010*agrarian 0.010*fishing 0.010*production 0.008*aid 0.008*farmer 0.007*product 0.007*oil 0.007*rural ... }
tProf *(diffFreq - ss)*	*t* = { 0.007*agriculture 0.007*sector 0.004*fishing 0.004*agrarian 0.004*production 0.004*aid 0.003*farmer 0.003*product 0.003*rural 0.002*oil ... }
tProf *(diffFreq - ls)*	*t* = { 0.008*agriculture 0.008*sector 0.005*fishing 0.005*agrarian 0.005*production 0.004*aid 0.003*farmer 0.003*product 0.003*rural 0.003*oil ... }
sProf *(diffFreq)*	*s* = { *0.216*"agricultural aid" 0.128*"agricultural policy" 0.099*"agricultural production" 0.098*"oily" 0.095*"food industry" 0.091*"fishing" 0.083*"oil" 0.075*"huelva province" ...* }

or a reference ontology, instead of using concepts with no structure, and this enables a much richer representation.

An example of this approach is described in [37], in which the authors build user profiles based on the user's browsing history by using concepts from the first three levels of the ODP ontology. Another example is [39], where the authors use weighted concept user profiles built on ontology-based semantic structures and metadata. They also build a context representation of the retrieval task, which is used to activate different parts of the user profile at runtime, thus matching the appropriate part of the user profile with the current retrieval task. In [22] the authors classify weighted thesaurus profiles of previous users into similar groups of users. When a new user arrives, the system recommends a profile to this user based on similarities with other users and it is then used for personalization or recommendation purposes. Finally, in [6] the authors show three different ways to use ODP: first, as a semantic support to find relations between concepts; secondly, to identify some ODP structure parts which are relevant to the user; and thirdly, for the user to directly choose the ODP concepts they are interested in. They then study how to use these three user profiles with query modification and reranking personalization techniques.

- *User profile update process.* Any interests and preferences in the user profile need to be constantly updated since they are dynamic and change over time [20,25]. An update process is therefore required so that the user profile remains accurate and up to date. This step is highly dependent on the two previous steps and should be considered in their design. For example, the use of implicit user information acquisition techniques highly facilitate this task. A static profile is only useful in very specific cases. Dynamic user profiles could be separated into long-term profiles (interests which define the user) and short-term profiles (more related to a specific information need). Both types of profiles will therefore evolve, but whereas the long-term one will be rather stable, the short-term one will be built and destroyed in a relatively short period of time.

Two examples where the authors try to combine both the long-term and short-term user profile approaches are Alipes [41], where three different vectors of weighted terms are used for each user interest: one for the long-term user profile, another for the short-term one (positive) and another for the short-term one (negative). The other example is detailed in [10], where the authors study how to learn long-term user interests by aggregating short-term user interests.

3. Term and subject based profiles

Both the profile information and its representation are important to produce good quality profiles, which are also essential to obtain good personalization results. For this reason, special care needs to be taken in the profile building process. In this article, we propose different approaches for the building process of our generic profiles and then compare their performance to obtain the best possible personalization results.

3.1. Profile building process

In this section we outline the design of the profile building process. For each of the proposed user profiles based either on terms or subjects, we shall explain its main characteristics and how to build and use them.

Term-based profiles (tProf): The first profile approach, based on the terms in the collection, can be considered as a *weighted keyword* profile, since the terms themselves are the items which represent user interests. These profiles are the easiest to build, but they need to have many terms to accurately define a user interest. These profiles are also less understandable for users than concept-based ones since their interests are much more easily mapped with concepts than with isolated terms. However, terms allow a more fine-grained representation of the collection content, see example in Table 1.

At first, we took the simple and common *tf*idf* weighting scheme [26] to build this kind of user profile. More specifically, each profile associated to an area of interest (committee) comprises the first *k* terms from documents of this area, ordered according to decreasing *tf*idf* weight values.

We soon realized, however, that user profiles following this weighting scheme included some terms not actually important or representative of any area of interest. Those terms were more connected with the documents format, such as *señor (sir)* to introduce a new speaker (see Table 1) or *gracias (thank you)* to express gratitude at the end of the speaker's speech, among others. In order to measure the impact of these specific collection stopwords, we decided to use two different stopword lists. We have used *ss (list of short stopwords)* to designate general purpose Spanish stopwords and *ls (list of long stopwords)* to designate long or extended stopwords with stopwords specific to our document collection, such as the previously mentioned *señor, gracias, etc*.

Subject-based profiles (sProf): This second approach, based on the initiative subjects, can be considered a *weighted concept* profile, since these subjects represent abstract topics of interest for the user instead of terms. They are represented as unstructured weighted concept vectors. The main advantages of general concept profiles are that they are robust to vocabulary variations and require less user feedback. These characteristics and the fact that the subjects are manually selected by experts in the document collection as the best initiative content representation lead us to consider that they would be a good resource to exploit.

In our opinion, the *tf*idf* tProf weighting scheme is not suitable for this kind of subject-based profile. Subjects are in fact some type of metadata-tag for the initiative content, usually from a controlled vocabulary (in our case from the EUROVOC thesaurus). Since this information is very concise and manually assigned by expert documentalists, the concept of *idf* (diminishing the influence of frequent items in the corpus) is meaningless for subjects.

We propose another weighting scheme for the profile items called ***diffFreq***, in order to avoid the two previous problems: 1) to be able to homogeneously build generic user profiles independently of the considered content (terms or subjects), and 2) to avoid to extend the language stopwords list to incorporate specific document collection stopwords. Additionally, as we shall see in Section 3.3, this new weighting scheme obtains better performance results than the *tf*idf* approach.

We now show how to select the elements of each profile type using the newly proposed weighting scheme *diffFreq*. Let *X* represent either a term in the case of *tProf* or a subject in the case of *sProf*, and let *Y* represent a profile. We then define $f^+(X, Y)$ as the frequency of *X* in documents belonging to any area(s) of interest which form the profile *Y*; $f^+(Y)$ is the number of elements (either terms for *tProf* or subjects for *sProf*) within *Y*; $f^-(X, Y)$ and $f^-(Y)$ are the frequency of *X* and the number of elements in documents outside the profile *Y*, respectively. We then define the *diffFreq* formula for the relevance of X with respect to Y as:

$$diffFreq(X, Y) = \frac{f^+(X, Y)}{f^+(Y)} - \frac{f^-(X, Y)}{f^-(Y)} \quad (1)$$

or in other words, the normalized frequency of *X* within *Y* minus the normalized frequency of *X* outside *Y*. If the final value is *diffFreq*(*X, Y*) ≤ 0, then *X* is more frequent outside *Y* than within and so it is not representative of *Y* and we will not consider it. However, if the final value is *diffFreq*(*X, Y*) > 0, this means that *X* represents *Y* to a certain degree and so we keep it. All the retained elements are sorted in decreasing order of relevance to form the final user profile.

Table 1 shows an example of the *tProf* and *sProf* learned profiles. For *tProf* we follow the *tf*idf* and *diffFreq* weighting schemes with both the *ss* and *ls* stopword lists. For *sProf* profiles we only show the *diffFreq* weighting scheme with the *ls* stopword list because of the reasons previously explained in the *sProf* definition.

Looking at Table 1 we can see how the *tProf* - *tf*idf* profiles change by using the *ss* or *ls* stopword lists. For example, the words *sir* and *andalusia* disappear and the words *oil* and *rural* appear, respectively. However the words following the new weighting scheme *diffFreq* are almost the same independently of the used stopwords list. The differences between the *tProf tf*idf - ls* and both *diffFreq* profiles are not observable in the table (apart from some word position inversions and slightly different weights), but there are some differences in further profile positions. However, the mere fact that *diffFreq* is able to remove the specific stopwords from the document collection already justifies its use. Moreover, as we shall see in Table 3, the performance is also better.

3.2. Evaluation framework

The evaluation framework, which will remain the same throughout this article, comprises the following components: 1) a document collection consisting of 658 committee sessions from the sixth and seventh Andalusian Parliament terms of office, marked up in XML (containing 432,575 retrievable structural units); 2) a heterogeneous set of 23 queries formulated by real users of the document collection; 3) Garnata [12] as the search engine and based on XML retrieval, firstly introduced by Chiaramella in 2001 [8]. XML retrieval is able to retrieve specific document parts that best meet the user's information needs, e.g. by returning a possibly relevant paragraph rather than the entire document as traditional IRSs do. This feature is particularly useful with large documents. Although we work with XML, the proposed user profiles do not depend on this documents format; 4) the relevance assessments were obtained from a conducted user study (ground truth). This user study involved 31 users, with a total of 126 evaluation triplets (user, query, profile), i.e. the relevance assessments provided by a given user, evaluating a given query under a given profile (considering each user chose the user profile closest to their personal interests and none of the user profile representations discussed in this article was shown to the user but rather a brief general description of the expected content); 5) the NDCG evaluation metric (normalized discounted cumulative gain) [17], with some special considerations due to the structured nature of the documents; and 6) the personalization techniques used are *NQE, HRR, NQE+m, HRR+m, CAS and CASor*, which represent a highly heterogeneous set of personalization techniques, with approaches from the three possible retrieval stages (in some cases mixing them) where personalization can be applied: before the search (e.g. query expansion - *NQE*), during the search (not yet often used, i.e. retrieval model modification - *NQE+m and HRR+m*) and after the search is performed (e.g. reranking - *HRR*). Additionally, we have also included two additional content-and-structure personalization techniques (*CAS and CASor*).

We shall now briefly explain these personalization techniques to enable a better understanding of their heterogeneity and main characteristics. For more details about these techniques or any other evaluation framework component see [13].

- *NQE (normalized query expansion)*: as its name suggests, this consists in adding the first *k* profile terms normalized by a factor called p_0 to the original query terms.
- *HRR (hard reranking)*: we perform two separate queries: the original query and the NQE query. We start a matching process between both lists of retrieved results and a matching occurs when the same result is found in both lists. The final personalized list of results is the original query results reordered according to the NQE matching order.
- *NQE+m* and *HRR+m* are the two previous techniques but using a modification of our Garnata search engine retrieval model. In short, this consists in differentiating between the original and expanded query terms in one step of the retrieval process to avoid the *query-drift* problem.
- *CAS* and *CASor*: an explanation of these personalization techniques is necessary as it will be required later in the article. These techniques use *content-and-structure* queries. These CAS queries allow us to make full use of the document structure, specifying in the query *what* we are looking for (the content), and *where* this should be located in the required documents (the structure).

In order to specify CAS queries, we have selected the widely used *NEXI* language [38]. The general form of a NEXI CAS query is *//A[B]//C[D]*. For example, the following CAS query attempts to retrieve chapters dealing with personalization and containing an INEX bibliography, in books with titles relating to information retrieval, where the chapter units are the target and the book units are the context:

```
//book[about(.//title, information retrieval)]
//chapter[about(., personalization) and about(.//bibliography, INEX)]
```

We shall transform the original query into a CAS query in such a way that its target part coincides with the original query and its context part contains the profile information. As the original query does not specify any structural restriction, we use the NEXI path wildcard operator "$*$" (meaning first or subsequent descendant) in the target part so that `//*[about(.,originalQueryTerms)]` is a CAS query equivalent to the original *content-only* query. For the context part of the query, we propose the use of the largest retrievable structural unit in the collection, `MaxUnit` (which is the least

Table 2
Examples of final *tProf* and *sProf* profiles using $k = 5$ and $p_0 = 0.66$ as user profile configuration parameters.

tProf *(tf*idf - ls)*	0.66*agriculture 0.639*sector 0.455*agrarian 0.452*fishing 0.450*production
tProf *(diffFreq - ls)*	0.66*agriculture 0.647*sector 0.401*fishing 0.399*agrarian 0.398*production
sProf *(diffFreq)*	*0.66*"agricultural aid" 0.390*"agricultural policy" 0.302*"agricultural production" 0.299*"oily" 0.291*"food industry"*

Table 3
Maximum (*max*), average (μ) and standard deviation (σ) performance values for the *tProf* profiles following the *tf*idf* and the *diffFreq* weighting schemes using *ss* (short stopwords) and *ls* (long stopwords).

		NQE	HRR	NQE+m	HRR+m	CAS	CASor
	max	0.608	0.624	0.637	0.645	0.629	0.601
tf∗idf	μ	0.517	0.569	0.597	0.593	0.611	0.546
(ss)	σ	0.076	0.048	0.040	0.046	0.011	0.035
	max	0.617	0.630	0.645	0.669	0.663	0.647
tf∗idf	μ	0.503	0.565	0.608	0.614	0.636	0.591
(ls)	σ	0.083	0.056	0.040	0.048	0.013	0.040
	max	0.617	0.626	0.661	0.676	0.667	0.643
diffFreq	μ	0.511	0.568	0.611	0.614	0.646	0.625
(ss)	σ	0.080	0.050	0.049	0.055	0.015	0.013
	max	0.615	0.626	0.666	0.681	0.671	0.650
diffFreq	μ	0.508	0.567	0.612	0.615	0.650	0.639
(ls)	σ	0.080	0.050	0.050	0.056	0.016	0.010

restrictive structural unit to hold the profile terms). The expanded CAS query would therefore be

$$//\mathtt{MaxUnit}[\mathtt{about}(.,\mathtt{profileTerms})]//*[\mathtt{about}(.,\mathtt{originalQueryTerms})]$$

Instead of using all the profile terms together, another option is to let each term comprise a different *about* clause, with all of these clauses being connected by the *or* operator. The reason for this modification is that in our case it is not necessary for all the profile terms to appear in the context part of a relevant structural unit. This new version of the expanded CAS query (CASor) is then

$$//\mathtt{MaxUnit}[\mathtt{about}(.,\mathtt{profileTerm1})\ \mathtt{or}\ \mathtt{about}(.,\mathtt{profileTerm2})\ \mathtt{or}\ldots\mathtt{or}\ \mathtt{about}(.,\mathtt{profileTermK})]//*[\mathtt{about}(.,\mathtt{originalQueryTerms})]$$

Before detailing our results, we shall explain how we have used these profiles according to the personalization techniques. All of the previous personalization techniques have an underlying common feature: in one way or another, all use an expanded query. There are therefore two main user profile parameters to use: a given number of expansion terms (*expTerms*) or subjects (*expSubj*) $k = 5, 10, 20, 40$ and a maximum weight normalization factor $p_0 = 0.33, 0.66, 0.99,$ which controls the importance of the expanded items with respect to the original query terms (weighted by 1.0). Once we have chosen the value of both parameters the k items normalized by p_0 will be the same for every personalization technique. The combination of both variables gives us a total number of $k(4)*p_0(3)$ different weighted terms or subject sets to provide to each personalization technique. Table 2 shows some examples of these final user profiles using *ls* stopwords from Table 1.

3.3. Results

This section shows the results of the different proposals for building user profiles based on terms *(tProf)* and subjects *(sProf)*, together with the conclusions derived from each approach. If not otherwise specified, each cell in all of the following results tables represents the average over the 126 evaluation triplets from the user study conducted for a given combination of expansion terms or subjects k, maximum normalization factor p_0 and a given personalization technique.

tProf results. We shall start by showing the results from the term-based user profiles. First, we wish to illustrate the differences in performance between the results obtained following the *tf*idf* and the new *diffFreq* weighting schemes with the two different stopword lists *ss* and *ls* (see Section 3.1). Table 3 shows the best (*max*), average (μ) and standard deviation (σ) performances of the *tf*idf* and the *diffFreq* weighting schemes, using *ss* and *ls* lists as stopwords for the term-based user profiles and under the previously mentioned evaluation framework.

As we can see in Table 3, the *tf*idf (ls)* approach improves *tf*idf (ss)* performance in most cases. More specifically, for the averaged six personalization techniques, *max* improves by 3.37%, μ improves by 2.41% and σ deteriorates by −10.06%. However, when *diffFreq (ls)* is compared with *diffFreq (ss)* there is only an improvement of 0.52% in *max*, 0.42% in μ and a deterioration of −0.32% in σ. These values show how the manual elimination of specific document collection stopwords (by using *ls* instead of *ss*) clearly improves *tf*idf* performance, while this is not the case for *diffFreq* which performs almost identically regardless of the stopword list used.

These results demonstrate how our newly developed weighting scheme *diffFreq* is able to eliminate almost all of the specific document collection stopwords by itself. Even if we compare *tf*idf (ls)*, which is the best *tf*idf* approach, with *diffFreq (ss)* the worst *diffFreq* approach, our new weighting scheme *diffFreq* still achieves 0.47% better *max* results and 1.65% better average results.

Combining the problems identified in Section 3.1, i.e. for the *tProf* profiles that follow the *tf*idf* weighting scheme unsuitable terms appear that need to be manually removed (or manually included in the stopword list used), with the unsuitable use of the *idf* component with subjects, and that the *diffFreq* weighting scheme has shown to perform better, it is clear why we have hereon chosen to follow the *diffFreq* weighting scheme.

Having demonstrated the suitability of the *diffFreq* weighting scheme for building generic profiles, Table 4 shows the NDCG average values considering term-based user profiles for all profile configuration parameters for each personalization technique under the given evaluation framework (best values in bold). We also show statistically significant differences between the best configuration values (in bold) and the other profile configurations for each personalization technique using a Student *t*-test at different significance levels.

From this table, we may draw the following main conclusions: firstly, personalized results are always better than the original non-personalized result (baseline), except for *NQE* with $k = 40$ and $p_0 = 0.99$; secondly, the combination of the best k and p_0 user profile parameters for each personalization technique (in bold) depends on the given personalization technique, with the highest values being obtained for those techniques that best avoid the *query-drift* problem (*NQE+m, HRR+m, CAS* and *CASor*), and relatively low values for those techniques that partially avoid this problem (*NQE* and *HRR*); thirdly, the absolute and averaged maximum performances are obtained by *HRR+m* and *CAS*, respectively, with $k = 40$

Table 4
NDCG averaged values for the *tProf* profiles (best values in bold).

k	p_0	NQE	HRR	NQE+m	HRR+m	CAS	CASor
5	0.33	**0.615**	0.621	0.516***	0.512***	0.618***	0.628***
5	0.66	0.572***	0.604**	0.601***	0.598***	0.644**	0.646
5	0.99	0.524***	0.582***	0.633***	0.640***	0.650*	0.647
10	0.33	0.605	**0.626**	0.538***	0.537***	0.629***	0.636*
10	0.66	0.524***	0.576***	0.626***	0.628***	0.655*	0.648
10	0.99	0.462***	0.542***	0.645**	0.656**	0.657*	0.650
20	0.33	0.575**	0.611	0.570***	0.564***	0.639***	0.630***
20	0.66	0.475***	0.553***	0.647*	0.652**	0.662*	0.645**
20	0.99	0.412***	0.507***	0.657*	0.670*	0.663*	**0.650**
40	0.33	0.541***	0.596*	0.587***	0.582***	0.645***	0.623***
40	0.66	0.428***	0.517***	0.658	0.665*	0.665	0.629***
40	0.99	0.365***	0.467***	**0.666**	**0.681**	**0.671**	0.632**
μ		0.508	0.567	0.612	0.615	**0.65**	0.639
σ		0.08	0.05	0.05	0.056	0.016	**0.01**
Baseline		0.381					

Note: |$^*p < .05$, $^{**}p < .01$, $^{***}p < .001$|.

Table 5
NDCG averaged values for the *sProf* profiles (best values in bold).

k	p_0	NQE	HRR	NQE+m	HRR+m	CAS	CASor
5	0.33	0.565	0.580	0.483***	0.478***	0.534	0.537**
5	0.66	0.518***	0.564*	0.547***	0.547***	0.531	0.536**
5	0.99	0.465***	0.534***	0.577***	0.578***	0.525	0.539**
10	0.33	**0.574**	**0.586**	0.495***	0.492***	0.537	0.556
10	0.66	0.514***	0.567	0.565***	0.568***	**0.542**	0.557
10	0.99	0.446***	0.527***	0.599**	0.608**	0.540	0.551**
20	0.33	0.561	0.575	0.504***	0.499***	0.519	0.560
20	0.66	0.490***	0.546*	0.576***	0.581***	0.533	0.560
20	0.99	0.419***	0.491***	0.613	0.619**	0.535	0.554*
40	0.33	0.538***	0.564*	0.510***	0.510***	0.483**	**0.565**
40	0.66	0.458***	0.515***	0.581***	0.590***	0.512*	0.563
40	0.99	0.389***	0.468***	**0.618**	**0.628**	0.520	0.554*
μ		0.495	0.543	0.556	**0.558**	0.526	0.553
σ		0.06	0.037	0.047	0.052	0.016	**0.01**
Baseline		0.381					

Note: |$^*p < .05$, $^{**}p < .01$, $^{***}p < .001$|.

and $p_0 = 0.99$ in the maximum case. The absolute maximum performance represents an improvement of 78.65% over the baseline.

sProf results. As explained in its definition, *sProf* user profiles only use the initiative subjects in their building process. In summary, each document initiative may have one or several subjects which have been manually selected by human documentalists as the best representation of the initiative content, while each subject may comprise one or several words. The results of applying these subject-based user profiles under the given evaluation framework are presented in Table 5 following the same Table 4 format.

However, before drawing certain conclusions from this table, we shall explain how subjects are really used by the different personalization techniques. A priori, this use should be exactly the same as when term-based user profiles are used *(tProf)* but because of the nature of the subjects, this is not exactly accurate.

Subject words are used as the expansion terms under the *NQE, HRR, NQE+m* and *HRR+m* personalization techniques. We should highlight two slight differences in relation to the *tProf* user profiles. The first difference is connected with the expansion process, i.e. although k subjects are still used in this expansion process since subjects comprise various different words (including some which may have been repeated between subjects), the total number of expansion words is unlikely to be exactly equal to k. The second difference is that although the subject words are obviously semantically related to their corresponding initiative content, the words themselves do not necessarily match the initiative terms. When these two differences are considered, they may result in a comparison between the results of the *tProf* and *sProf* profiles (for each configuration of the k and p_0 parameters of the user profiles) that would not have been made a priori under the same conditions. Nevertheless, the trends of these results and the general conclusions arising from them are still valid and are in fact very similar. In this respect, we can see how the *sProf* profile results shown in Table 5 for the first four personalization techniques are lower than those for the *tProf* profiles in Table 4. The maximum values, averages and even standard deviation trends, on the other hand, are quite similar.

The use of the *CAS* and *CASor* personalization techniques is also a bit different from the *tProf* user profile approach. If we consult Section 3.2 *CAS* and *CASor* personalization techniques underlying CAS queries, now that we are using subjects instead of terms these will be transformed into the following expressions, respectively:

```
//MaxUnit[about(.//subjects,profileSubjects)]
//*[about(.,originalQueryTerms)]
   //MaxUnit[about(.//subjects,profileSubject1)
or about(.//subjects,profileSubject2)
   or... or about(.//subjects,profileSubjectK)]
//*[about(.,originalQueryTerms)]
```

It can be seen how in the new *CAS* queries using the *sProf* user profiles, the previous *profileTerms* are replaced by *profileSubjects*. However, as the reader may observe, these profile subjects are now only searched in the initiative-associated *subject* tags where the subjects are located and not in the entire *MaxUnit* content. Therefore, the new *CAS* queries for *sProf* profiles search for the original query terms anywhere in the document but only the results with initiatives where *profileSubjects* appear will actually be retrieved (the higher the number of *profileSubjects*, the better). On the other hand, *CASor* relaxes this requirement, i.e. the number of *profileSubjects* required.

We have decided to allow *sProf* CAS approaches to only search for subjects in the initiative *subject* tags so as to avoid the previous *sProf NQE,HRR,NQE+m,HRR+m* observed unmatching problem between the initiative-assigned subjects and their content terms (although both are semantically related).

Table 5 with the *sProf* user profiles shows a considerably worse performance for *CAS* and *CASor* personalization techniques in comparison with their Table 4 *tProf* counterparts. As a result, we made substantial efforts to implement and test many design variations for both CAS approaches, including not to propagate to the *MaxUnit* but to the *initiative* and other structural units but with no success. We even replicated the behaviour of the *tProf* CAS approach, i.e. not to search for the subjects in the *subject* tags but anywhere in the content, which may be considered similar behaviour to the *sProf NQE, HRR,NQE+m* and *HRR+m* approaches. Although there was a slight improvement in the results, this would mean considering subjects as simple terms, and since CAS queries enable specific places to be searched, we wish to treat subjects as subjects rather than terms.

4. Subjects and terms working together

The use of subjects in user profiles seems to be challenging as demonstrated by the low *sProf* profile performance, particularly for the CAS and CASor personalization techniques. However, since we still believe in the use of subjects for personalization purposes and in order to avoid semantic problems or other problems arising from the use of subjects as query expansion keywords, we developed a new personalization technique. This technique combined the use of subjects and terms in an attempt to avoid or at least diminish such problems and to obtain the best of both kinds of content.

4.1. CASmix

We have called this newly developed personalization technique ***CASmix***. As its name suggests, this is a hybrid CAS personalization technique which combines subjects from the *sProf* profiles and terms from the *tProf* profiles as follows: the initiative-associated *subjects* tag are searched for the *sProf* subjects and the *MaxUnit* content is searched for the *tProf* terms. The underlying CAS query is therefore as follows:

```
//MaxUnit[about(.//subjects,profileSubjects)
or about(.,profileTerms)]
//*[about(.,originalQueryTerms)]
```

The idea behind this new approach is to try to make the most of subjects and terms in combination instead of separately as we have until now. If we look at the performance of this new technique in Table 6 and compare it with all the other techniques in Table 5, we can see that *CASmix* obtains the best maximum and average performances. We can therefore conclude that we do not achieve particularly good results if we only use subjects as the profile information, but performance is much better if we use subjects together with terms rather than subjects alone.

Furthermore, if we compare the results in Tables 5 and 4, each *sProf* technique performs worse than the corresponding *tProf* technique. This suggests that it is better to use terms instead of subjects to build the user profiles. However, if we consider that the newly developed *CASmix* personalization technique performs better than both *tProf* CAS approaches, not only for the highest user profile configuration parameters performance but also on average, it seems that the use of subjects together with terms is a better approach than only using terms. The performance obtained by the *CASmix* approach is actually the highest of the entire set of personalization techniques for both *tProf* and *sProf* profiles (except for the highest *tProf HRR+m* configuration) and the overall best average result.

Table 6
NDCG averaged values for the newly developed *CASmix* personalization technique.

k	p_0	CASmix
5	0.33	0.637**
5	0.66	0.641**
5	0.99	0.642**
10	0.33	0.657*
10	0.66	0.663
10	0.99	0.663
20	0.33	0.663**
20	0.66	0.665*
20	0.99	0.667
40	0.33	0.666
40	0.66	**0.672**
40	0.99	0.671
μ		0.659
σ		0.012
Baseline		0.381

Note: |*p < .05, **p < .01, ***p < .001|.

4.2. Profiles based on subjects and terms

In view of these results, it seems that the combination of subjects and terms in the personalization process is a good strategy to obtain good personalization results and is better than using any of them individually. We have therefore given it a further twist by using the same idea within the profile itself.

We have developed the *stProf* profile, which joins these two elements in the same profile. More specifically, *stProf* profiles will comprise those subjects which best represent user profile interests in a first level, with each of these subjects having a set of their most representative terms in a second level. This kind of user profile is therefore a hybrid approach between the *weighted concept* and *weighted keyword* profile representations, enabling concept abstraction to be kept yet enriched by the fine-grained contribution of the terms.

4.2.1. Building process.

In order to calculate the *stProf* profile subjects and terms, we follow the same *diffFreq* formula (see Eq. 1). The subjects and their weights in the first level of these new *stProf* profiles will be exactly the same subjects and weights as those in the *sProf* profiles. This is because the subjects in the first level of this *stProf* profile represent the profile itself. However, in order to calculate the second level terms, which represent each first level subject (in this case the subject could be considered as the *"profile"*), the formula is interpreted as follows: in this case X represents a term and Y represents a subject, and in this case $f^+(X, Y)$ is the frequency of X in initiatives classified by the subject Y and $f^+(Y)$ is the total number of terms in these initiatives; $f^-(X, Y)$ and $f^-(Y)$ are, respectively, the frequency of X and the number of terms in documents outside the initiatives classified by the subject Y.

As the reader may observe, following the previous calculation methodology, the same subject will have exactly the same set of associated terms independently of the profile being considered. We

Table 7
Examples of the associated terms (unstemmed, with fewer decimals and translated into English) for the subject "*huelva province*" under different generic profiles.

diffFreq	*agriculture and livestock* *environment*	0.008*huelva 0.003*zone 0.003*province 0.003*sector 0.002*fishing 0.008*huelva 0.003*zone 0.003*province 0.003*sector 0.002*fishing
diffFreq alternative	*agriculture and livestock* *environment*	0.010*sector 0.008*huelva 0.008*fishing 0.006*zone 0.004*fishery 0.008*environment 0.007*medium 0.006*huelva 0.005*fire 0.005*residue

Table 8
Example of the *stProf* profiles, for the 'agriculture and livestock' area of interest (unstemmed, with fewer decimals and translated into English) following the *diffFreq* weighting scheme.

stProf	s_1 =*0.216*"agricultural aid"* s_2 =*0.128*"agricultural policy"* ⋮	t_{s1} = {0.008*aid 0.007*sector 0.006*agriculture 0.006*farmer ... } t_{s2} = {0.010*agriculture 0.007*agrarian 0.007*production ... } ⋮

have also tested other ways of calculating these associated terms, in which the same subject would have a different, more specific set of associated terms depending on the profile being considered. Table 7 shows examples for *diffFreq* and one of these alternative approaches.

The evaluation results of these alternatives, however, where the subject-associated terms will also depend on the given profile are slightly worse than when the original *diffFreq* is used. In the following sections of this article, therefore, we have finally decided to only use the original *diffFreq* approach for *stProf* profiles, where the same subject will have exactly the same set of associated terms regardless of the profile considered.

Table 8 shows an example of *stProf* learned profiles following the selected *diffFreq* weighting scheme.

4.2.2. *How to use the* stProf *profiles.*

The use of *stProf* is somewhat more complicated than the use of the *tProf* or *sProf* profiles. In principle, the process should be to obtain the first *expSubj* profile subjects, and for each of these subjects to obtain the first *expTerms* terms. Each term weight will be multiplied by its corresponding subject weight. The terms will therefore be those eventually used by the personalization techniques and will already include in their weights the influence of the subject they represent.

This process does, however, have a problem: when joining the different terms associated to different subjects, some of these terms are repeated (several subjects have terms in common, such as *agriculture* in the example in Table 8). Since there is little point in having repeated terms with different weights, the following approaches are considered in order to fix the weights of these terms:

1. **stProf_add** *(add weights)*: collapse the repeated terms into one, with a weight equal to the addition of the individual weights.
2. **stProf_max** *(maximum among weights)*: we only keep the repeated term with the highest weight and remove all other repeated terms.
3. **stProf_addFill** *(add weights, filling terms)*: same as *stProf_add*, but each time a term is deleted from a subject, the next one in the list of terms of this subject is included until there are exactly *expTerms* terms for each subject.
4. **stProf_maxFill** *(maximum among weights, filling terms)*: same as *stProf_addFill*, but using the maximum instead of the sum.

The first two approaches mean that we do not always obtain the same number of terms for the personalization techniques, as occurs with the last two approaches. It should be noted that in the last two approaches the filling process should start from the last *expSubj* subject, since we want more information from the most representative subjects of the profile, i.e. the first subjects in the ranking. At the end of this process, the final terms will also be normalized with a maximum normalization value p_0.

We can see examples of these four different *stProf_** user profile approaches in Table 9, based on the *stProf* user profile example in Table 8. Some of the characteristics of these approaches are: the weight of each approach's first term is equal to 0.66 (p_0); while the first two approaches have five terms instead of six, because the 'agriculture' term is repeated in both *expSubj* terms, the last two approaches have six terms, since terms are added until there are *expSubj*∗*expTerms* terms in total; since the term 'agriculture' is repeated, the sum of both weights places it as the first term in the **add** approaches, while 'aid' is the first term in the **max** approaches; the term 'farmer' which belongs to the subject "agricultural aid" is the term added by the **Fill** approaches (considering that the addition process starts from the last *expSubj* subject).

4.2.3. *Results.*

For *stProf* profiles the combination of the different possible values of *expSubj, expTerms* and p_0 variables gives us a total number of *expSubj*(4)∗*expTerms*(3)∗p_0(3) different weighted term sets to provide to each personalization technique, as it can be seen in Table 10.

In principle, as the *stProf* profiles actually use terms rather than subjects in the expansion process, they should at least partially solve the *sProf* profile problems: the same number of expansion terms will be used (this is only true in the **Fill** approaches) and the expansion terms will match the document content being not simply semantically related to it, as occurred when subjects were used. Meanwhile, as the expansion term weights have already been multiplied by their corresponding subject weight, they already have the influence of the subjects, which as we have already seen in the *CASmix* personalization technique contribute in some way. We shall now see whether this assumption is reinforced by the *stProf* profile results.

The evaluation framework is exactly the same as the one in Section 3.2, with the only difference being that with *stProf* profiles k represents the number of subjects and $l = 1, 5, 10$ represents the number of expansion terms for each subject.

If we were to show the four *stProf_*∗ profile approaches, we would be required to compile four tables (one per page) consisting of 36 rows (user profile configurations) and 6 columns (personalization techniques) cells. For conciseness, we shall therefore only show the *stProf_maxFill* results (the best of the four approaches) in Table 10. However, we attach figures for the four approaches results and some conclusions about their comparison in Appendix A. It should be noted that the *CASor* personalization technique with the user profile configuration parameter $k = 40$ has no available results since the execution time (particularly for $l = 10$) is unacceptable even for experimental purposes.

The following main conclusions can be drawn from Table 10: 1) *HRR* always performs better than *NQE* and both obtain worse results with higher *k, l* and p_0 variables values because of the

Table 9
Final *stProf* user profile using $expSubj = 2$, $expTerms = 3$ (for reasons of clarity and conciseness) and $p_0 = 0.66$.

stProf_add	0.66*agriculture 0.440*aid 0.374*sector 0.237*agrarian 0.222*production
stProf_max	0.66*aid 0.560*sector 0.525*agriculture 0.356*agrarian 0.333*production
stProf_addFill	0.66*agriculture 0.440*aid 0.374*sector 0.311*farmer 0.237*agrarian 0.222*production
stProf_maxFill	0.66*aid 0.560*sector 0.525*agriculture 0.466*farmer 0.356*agrarian 0.333*production

Table 10
NDCG averaged values for the *stProf_maxFill* profiles (best values in bold).

k	l	p_0	NQE	HRR	NQE+m	HRR+m	CAS	CAS-or
5	1	0.33	0.586	0.584**	0.488***	0.486***	0.594***	0.606***
5	1	0.66	0.557***	0.581***	0.555***	0.559***	0.624**	0.622***
5	1	0.99	0.503***	0.554***	0.599***	0.608***	0.630*	0.626**
5	5	0.33	0.543***	0.578***	0.555***	0.555***	0.624**	0.613***
5	5	0.66	0.442***	0.517***	0.626***	0.629***	0.642	0.623***
5	5	0.99	0.375***	0.474***	0.640**	0.646***	0.645	0.627***
5	10	0.33	0.518***	0.568***	0.569***	0.566***	0.625**	0.601***
5	10	0.66	0.399***	0.488***	0.635***	0.637***	0.641	0.608***
5	10	0.99	0.340***	0.449***	0.647**	0.658***	0.648	0.609***
10	1	0.33	**0.603**	**0.619**	0.522***	0.521***	0.612***	0.621***
10	1	0.66	0.531***	0.587**	0.602***	0.604***	0.635*	0.642
10	1	0.99	0.465***	0.540***	0.637***	0.646***	0.641	0.645
10	5	0.33	0.521***	0.566***	0.573***	0.570***	0.623**	0.604***
10	5	0.66	0.406***	0.491***	0.641**	0.647***	0.645	0.610***
10	5	0.99	0.339***	0.435***	0.655*	0.664***	0.648	0.613***
10	10	0.33	0.481***	0.540***	0.583***	0.580***	0.618**	0.588***
10	10	0.66	0.367***	0.456***	0.644**	0.651***	0.642	0.593***
10	10	0.99	0.305***	0.405***	0.654*	0.668***	0.648	0.594***
20	1	0.33	0.558***	0.592**	0.550***	0.550***	0.619***	0.629***
20	1	0.66	0.470***	0.524***	0.628***	0.631***	0.641	0.643***
20	1	0.99	0.397***	0.480***	0.653**	0.658***	0.647	**0.648**
20	5	0.33	0.475***	0.525***	0.587***	0.582***	0.620**	0.592***
20	5	0.66	0.367***	0.450***	0.649*	0.659**	0.645	0.597***
20	5	0.99	0.309***	0.399***	0.658*	0.671***	0.648	0.598***
20	10	0.33	0.446***	0.501***	0.594***	0.591***	0.613***	0.578***
20	10	0.66	0.336***	0.423***	0.655	0.665**	0.634*	0.582***
20	10	0.99	0.290***	0.392***	0.667	0.683*	0.646	0.582***
40	1	0.33	0.541***	0.580**	0.571***	0.569***	0.620***	–
40	1	0.66	0.427***	0.500***	0.645**	0.648***	0.642*	–
40	1	0.99	0.366***	0.460***	0.664	0.671**	**0.649**	–
40	5	0.33	0.445***	0.499***	0.596***	0.591***	0.608***	–
40	5	0.66	0.335***	0.422***	0.657	0.664**	0.636	–
40	5	0.99	0.290***	0.392***	0.664	0.683*	0.646	–
40	10	0.33	0.418***	0.477***	0.598***	0.595***	0.595***	–
40	10	0.66	0.314***	0.407***	0.659	0.670*	0.624**	–
40	10	0.99	0.276***	0.376***	**0.671**	**0.687**	0.634	–
μ			0.426	0.495	0.616	0.621	**0.632**	0.611
σ			0.094	0.069	0.046	0.051	**0.015**	0.02
Baseline			0.381					

Note: |*p < .05, **p < .01, ***p < .001|.

query-drift problem; 2) *NQE+m* and *HRR+m* obtain similar results with the latter being slightly better. Unlike the corresponding personalization techniques without *+m*, these approaches obtain better results with higher *k, l* and p_0 variable values. This is clear proof that these techniques solve the *query-drift* problem; 3) *CAS* and *CASor* obtain similar outputs and are also similar to the previous *+m* techniques; and 4) the maximum absolute performance 0.687 is obtained by *HRR+m* with $k = 40$, $l = 10$ and $p_0 = 0.99$ representing an improvement of 80.17% over the baseline.

5. Comparing all the profiles

In this section, we shall amalgamate and summarize all of the previously presented results for the six different developed user profile approaches. Since the newly developed *CASmix* personalization technique uses the *tProf* and *sProf* profile terms and subjects, we combine its results with the *sProf* profile results so as not to repeat the same value in the *tProf* profile results and because its development arose from and was based on the *sProf* profiles with the aid of terms.

Tables 11 and 12 show all the proposed user profiles results. As these are summary tables, *stProf* profiles others than *stProf_maxFill* are also included. For a deeper information check Appendix A.

The *general main conclusion* from Table 11 is that generic profiles personalization always (except in exceptional cases) helps the user to find relevant information faster and easier. Additionally, if we also carefully select the proper configuration parameters for any of the proposed user profiles and personalization technique, we always obtain a relatively good improvement in personalization with respect to the non-personalized IRS performance (NDCG = 0.381) ranging from 53.44% to 80.17%.

There are other three main conclusions to be drawn from Table 11: 1) the best personalization techniques in maximum, average and standard deviation NDCG values are clearly *HRR+m, CAS* and *CASor*, respectively; 2) the best user profile approach for maximum performance values is *tProf* with the exception of *NQE+m* and *HRR+m* where the *stProf_maxFill* profile is better. In the case of average and standard deviation values, there is more variability. For *NQE+m* and *HRR+m*, however, the *stProf_maxFill* profile is

Table 11
NDCG maximum, average (μ) and std. (σ) performance values for the six developed user profile approaches under the evaluation framework. Original (non-personalized) NDCG value: 0.381. Cells in bold shows the best user profile approach for each personalization technique, and '+' character shows the best personalization technique for a given user profile approach.

		NQE	HRR	NQE+m	HRR+m	CAS	CASor	CASmix
max	*tProf*	**0.615**	**0.626**	0.666	0.681$^+$	**0.671**	**0.650**	–
	sProf	0.574	0.586	0.618	0.628	0.542	0.565	**0.672**$^+$
	stProf_add	0.593	0.614	0.655	0.666$^+$	0.652	0.643	–
	stProf_max	0.585	0.609	0.666	0.674$^+$	0.654	0.647	–
	stProf_addFill	0.611	0.619	0.660	0.675$^+$	0.654	0.646	–
	stProf_maxFill	0.603	0.619	**0.671**	**0.687**$^+$	0.649	0.648	–
μ	*tProf*	0.508	**0.567**	0.612	0.615	**0.650**$^+$	**0.639**	–
	sProf	0.495	0.543	0.556	0.558	0.526	0.553	**0.659**$^+$
	stProf_add	**0.514**	0.560	0.578	0.581	0.629$^+$	0.623	–
	stProf_max	0.481	0.533	0.591	0.597	0.630$^+$	0.625	–
	stProf_addFill	0.458	0.524	0.605	0.607	0.635$^+$	0.616	–
	stProf_maxFill	0.426	0.495	**0.616**	**0.621**	0.632$^+$	0.611	–
σ	*tProf*	0.080	0.050	0.050	0.056	0.016	0.010$^+$	–
	sProf	**0.060**	**0.037**	0.047	0.052	0.016	0.010$^+$	**0.012**
	stProf_add	0.069	0.042	0.054	0.058	0.019	**0.010**$^+$	–
	stProf_max	0.081	0.057	0.054	0.057	0.018	0.011$^+$	–
	stProf_addFill	0.093	0.064	0.047	0.052	**0.015**	0.014$^+$	–
	stProf_maxFill	0.094	0.069	**0.046**	**0.051**	0.015$^+$	0.020	–

Table 12
User profile parameters $k[-l]-p_0$ configuration for each maximum NDCG personalization technique-user profile performance, with '$*$' and '+' characters meaning the same as in Table 11.

	NQE	HRR	NQE+m	HRR+m	CAS	CASor	CASmix
tProf	05-0.33*	10-0.33*	40-0.99	40-0.99$^+$	40-0.99*	20-0.99*	–
sProf	10-0.33	10-0.33	40-0.99	40-0.99	10-0.66	40-0.33	40-0.66^{+*}
stProf_add	40-01-0.33	10-01-0.66	40-10-0.99	40-10-0.99$^+$	20-05-0.99	20-01-0.99	–
stProf_max	10-01-0.33	10-01-0.66	40-10-0.99	40-10-0.99$^+$	20-01-0.99	20-01-0.99	–
stProf_addFill	10-01-0.33	10-01-0.33	40-10-0.99	40-10-0.99$^+$	20-01-0.99	20-01-0.99	–
stProf_maxFill	10-01-0.33	10-01-0.33	40-10-0.99*	40-10-0.99^{+*}	40-01-0.99	20-01-0.99	–

Table 13
General NDCG maximum (*max*), average (μ) and deviation (σ) values for each of the six proposed user profile approaches.

	tProf	sProf	stProf_add	stProf_max	stProf_addFill	stProf_maxFill
max	**0.652**	0.598	0.637	0.639	0.644	0.646
μ	**0.598**	0.556	0.581	0.576	0.574	0.567
σ	0.044	**0.034**	0.042	0.046	0.047	0.049

again the best; and 3) the newly developed personalization technique *CASmix* clearly outperforms *sProf* profiles and also *tProf* profiles in most cases. Additionally, when compared to all the other approaches, it achieves a considerably high maximum NDCG performance after some of the *HRR+m* configurations, but by far the highest average value and one of the lowest standard deviations of all approaches.

In view of these conclusions, we may assume that generally the best user profile approach to use is the simpler *tProf* rather than the slightly more complicated *stProf_maxFill* which only achieves an improvement of 0.85% over the previous maximum performance obtained under the *tProf* user profile.

Table 12 shows which user profile configuration, i.e. $k[-l]-p_0$ parameters, maximizes performance. If we focus on the *best personalization technique* for each user profile approach ('+' character), the user profile configuration that maximizes performance is clearly 40-[10]-0.99, except for *CASmix* where it is 40-0.66. *HRR+m* is the best personalization technique for every profile except the *sProf* profile where *CASmix* is again the best. However, if we focus on the *best user profile* approach for every personalization technique ('$*$' character), the user profile configuration that maximizes performance comprises low values such as $k = 5, 10$ and $p_0 = 0.33$ for the *NQE* and *HRR* techniques and high values such as 40-[10]-0.99 for every other personalization technique with the exception of *CASor* and *CASmix* which have values of 20-0.99 and 40-0.66, respectively.

As we can see, these user profile configuration values basically depend on the given personalization technique rather than on the kind of user profile. Once again, these conclusions verify that personalization techniques which solve the well-known *query-drift* problem achieve their maximum performance when using the highest values of the user profile configuration parameters, i.e. using the largest amount of information.

Table 13 , finally, shows the general NDCG maximum (*max*), average (μ) and deviation (σ) values for each of the six proposed user profile approaches, i.e. this table shows the general average expected results for any given personalization technique for the six different user profile approaches. Each cell represents the average value for each personalization technique (row) from Table 11 for a given profile approach.

It is possible to observe how the maximum and minimum *max* performances are achieved by the *tProf* and *sProf* user profiles, respectively, while the *stProf* profiles obtain relatively good values which increase as we move to the right of the table. It is also apparent how the highest average (μ) value is achieved by the *tProf* approach, with a relatively low deviation (σ) value. Meanwhile, the

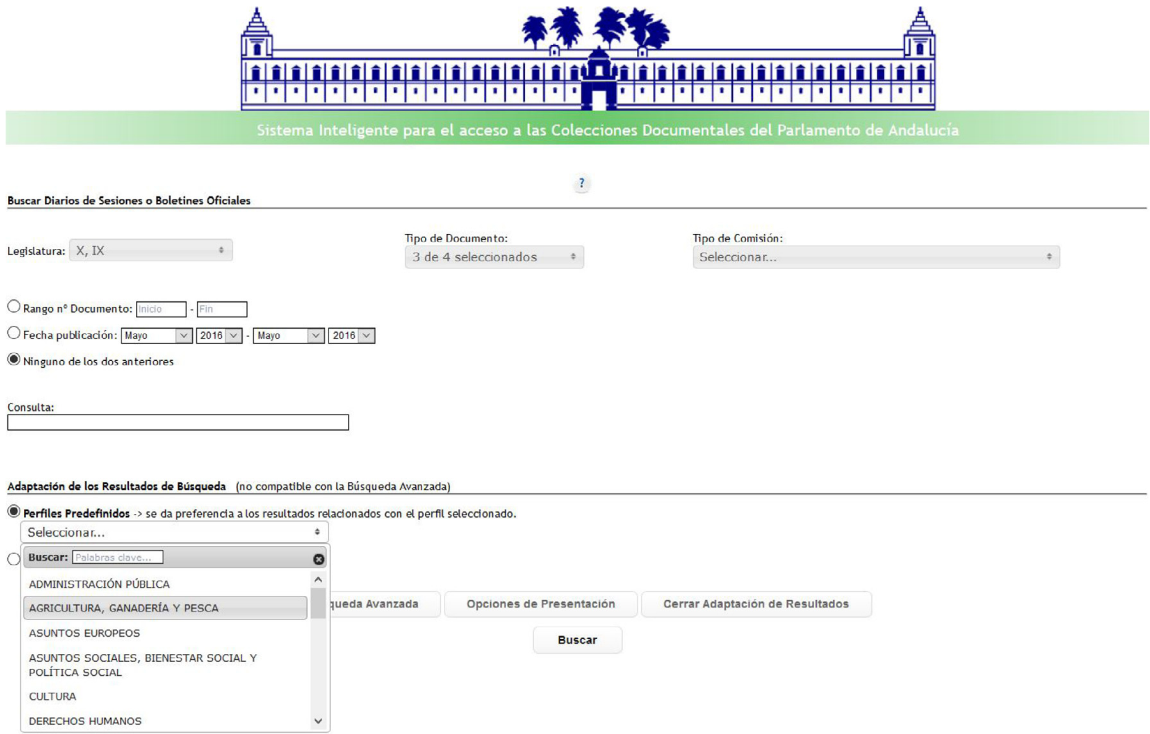

Fig. 2. SEDA-personalized IRS interface.

lowest deviation value is achieved by the *sProf* approach, but with a much lower average value than *tProf*. Considering the four *stProf* approaches, we can observe a gradual decrease and increase in the average and deviation values, respectively, following the order of these profiles in the table. This situation indicates that within these user profiles, the further to the right they are in the table, the more disparate personalization results (higher and lower) they achieve and so more attention needs to be paid when selecting the right user profile configuration. This conclusion is also confirmed by the fact that the maximum experimental evaluation performance is achieved with the *stProf_maxFill* approach.

In view of all these results, it might appear that the *tProf* profile is the best alternative, but this is not necessarily the case. From the user's point of view and when not particularly small profiles are considered, an *stProf* profile is much easier to understand than a *tProf* profile since abstract concepts contain more semantics than isolated terms. It is also true that the *stProf* profile with two levels (concepts and terms) could be exploited by a given personalization technique to improve its performance, e.g. easily selecting parts of the user profile which best match the query (particularly helpful for heterogeneous profiles). Depending on the application and personalization technique used, therefore, a trade-off decision must be reached between pure performance and greater user profile expressiveness.

6. General conclusions and future work

Since user profiles are very important for personalization, in this article we have presented six different generic user profile representations based on content. Although these generic profiles do not represent real users, and this is in fact their main disadvantage, they are suitable for representing user interests and do have many other advantages. For example, they are perfect for enabling personalization in privacy-constrained environments, since they do not collect any personal information. As a result, they do not place any burden on the user nor require any complex user gathering information process. They are also easier and less expensive to maintain, since there will be only a few of them (as many as collection categories, which are much fewer than possible IRS users) and they only need to be updated when the content suffer any modification (e.g. updating them in the small hours of the morning). Two further advantages are that they can be stored on the server, thereby reducing network traffic and more importantly, without the need to send personal information with the involved risks through the network, and that they could even be used as the first version of a real user profile ('cold-start').

In this article, we have developed a new way to build user profiles based on terms (*tProf*) using a new weighting scheme called *diffFreq* which improves the classical *tf***idf* approach, at least in this category-based generic profiles, and it is also compatible with other content sources (e.g. subjects or categories). We have then presented user profiles based on subjects considered as concepts (*sProf*), which are manually assigned by documentalists from a thesaurus to the document initiatives. This second user profiles performed relatively poorly in comparison with *tProf* profiles. We therefore attempted to improve their results with different approximations until we finally developed a new personalization technique (*CASmix*), which uses both subjects and terms obtaining quite good performance results. Finally, and inspired by *CASmix*, we have proposed a new hybrid profile approach (*stProf*) based on the two previous profile approaches (and with four variations) with a two-level representation, where the first level is represented by subjects and the second level by the terms representing these subjects.

We have performed a comprehensive evaluation experimentation which include six different personalization techniques (seven with CASmix) and a wide range of user profile configurations for each of the proposed user profile approaches. We have obtained very good personalization results which revealed an improvement of up to *80.17%* with respect to the original non-personalized IRS performance. Additionally, we have demonstrated that the use of a simple term-based user profile is normally enough to obtain good

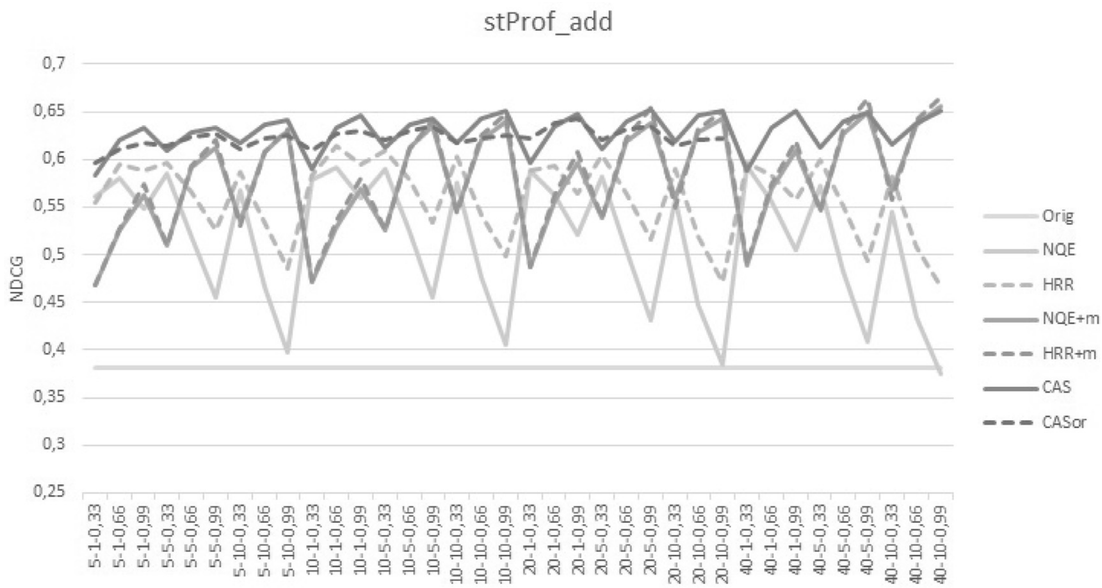

Fig. A.3. stProf_add results.

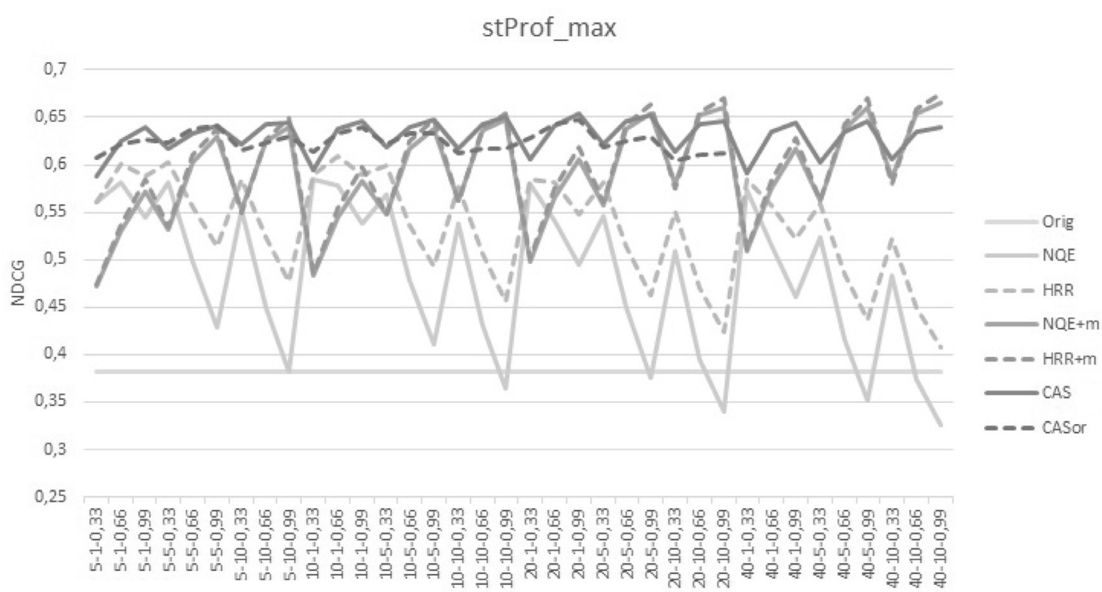

Fig. A.4. stProf_max results.

personalized results. Even so, having a user profile with a certain structure and abstract concepts may not only help users to better understand their own profiles, but also enable personalization techniques to exploit this richer representation. In fact, these user profiles with concepts are particularly suitable for certain IR subfields such as, for example, multilingual IR [16].

We are particularly proud to have recently integrated these generic profiles (for the time being the term-based approach) into the live, privacy-constrained *Seda* environment of the Andalusian Parliament. Fig. 2 shows the IRS interface, where the user can obtain personalized results for a given query by selecting any of the predefined profiles. Feedback from these IRS users about this new personalization feature has been extremely positive, since although the system does not ask for any personal information, it is now able to satisfy their information needs more easily and faster.

As *future work*, we would like to include in some way additional information to the user profiles, such as localization or temporal information. We would also like to develop some personalization techniques in order to better exploit the hierarchy of the *stProf* user profiles. Another possible alternative would be the possibility of dynamically choosing whether to use personalization and of selecting certain parts of the profiles according to user context or query characteristics. And last, but not less important, we would definitely like to use these proposed generic profiles to personalize other real environments with privacy restrictions, and of course, continue improving our work with the Andalusian Parliament.

Acknowledgements

This work was supported by the Spanish Andalusian "Consejería de Innovación, Ciencia y Empresa" postdoctoral phase of project P09-TIC-4526, the Spanish "Ministerio de Economía y Competitividad" project TIN2013-42741-P and the European Regional Development Fund (ERDF-FEDER).

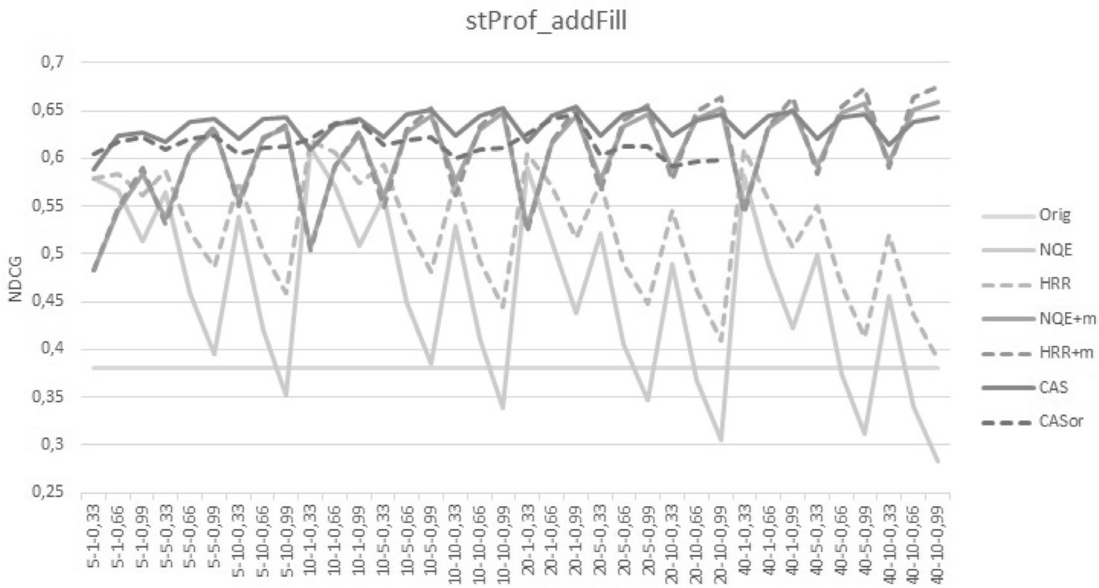

Fig. A.5. stProf_addFill results.

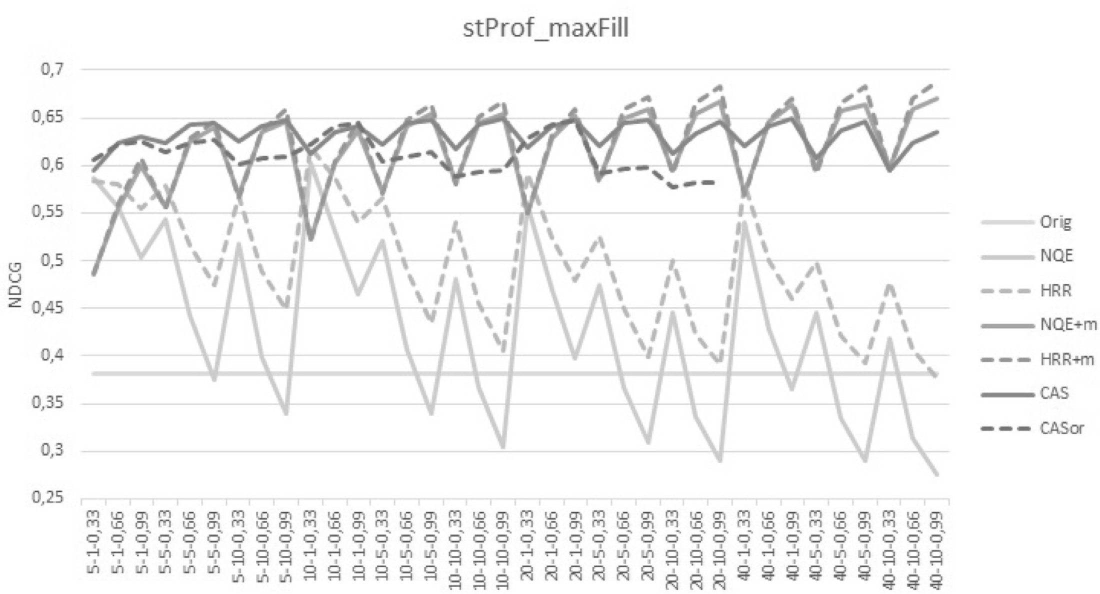

Fig. A.6. stProf_maxFill results.

Appendix A

In this appendix, we show the results of the four *stProf_∗* profile approaches (see Section 4.2) with Figs. A.3, A.4, A.5 and A.6.

These are the main drawn conclusions from the previous *stProf_∗* profile approaches:

- *stProf_max* and *stProf_maxFill* always obtain higher maximum and averaged performance results than *stProf_add* and *stProf_addFill*, respectively, for the *+m* personalization techniques, while obtaining lower results under the *NQE* and *HRR* approaches (hereafter denoted as *base* personalization techniques). This fact shows how **max** approaches are more suitable for personalization techniques which best avoid the *query-drift* problem, while **add** approaches are more suitable for personalization techniques which partially avoid this problem.
- Both **Fill** approaches always obtain higher maximum results than their corresponding *'noFill'* approaches, except for *CAS* in *stProf_maxFill vs. stProf_max*, i.e. in 11 out of 12 cases. When we focus on the average results, however, both previous **Fill** versus *'noFill'* profile comparisons obtain lower performance results considering the *base* personalization techniques, but higher values considering the *+m* techniques. This means that by carefully selecting the user profile configuration parameters, **Fill** user profiles are always better. If we are not sure about the suitability of these profile parameters, we can still trust **Fill** user profiles for *+m* personalization techniques, but it would be better to use *'noFill'* user profiles for *base* personalization techniques.
- *CAS* approaches obtain very similar results between the four different *stProf* user profiles and have low standard deviation values. There is therefore not much difference between using any of the four user profile approaches.
- Figs. A.3, A.4, A.5 and A.6 reveal how more disparate personalization results (higher and lower) are obtained between the

first and the last figure. More attention must therefore be paid when selecting the right user profile configuration.

- *stProf_maxFill-HRR+m* and *stProf_addFill-CAS* user profile-personalization techniques obtain the absolute and averaged maximum performances, respectively, with $k = 40$, $l = 10$ and $p_0 = 0.99$ in the maximum value. The absolute maximum performance represents an improvement of 80.17% over the baseline.

References

[1] J.-w. Ahn, P. Brusilovsky, J. Grady, D. He, S.Y. Syn, Open user profiles for adaptive news systems: help or harm? in: Proceedings of the 16th International Conference on World Wide Web, ACM, 2007, pp. 11–20.
[2] J.-W. Ahn, P. Brusilovsky, D. He, J. Grady, Q. Li, Personalized web exploration with task models, in: Proceedings of the 17th International Conference on World Wide Web, ACM, 2008, pp. 1–10.
[3] M.Y.H. Al-Shamri, User profiling approaches for demographic recommender systems, Knowl. Based Syst. 100 (2016) 175–187.
[4] J. Allan, J. Aslam, N. Belkin, C. Buckley, J. Callan, B. Croft, S. Dumais, N. Fuhr, D. Harman, D.J. Harper, et al., Challenges in information retrieval and language modeling: report of a workshop held at the center for intelligent information retrieval, university of massachusetts amherst, september 2002, in: ACM SIGIR Forum, 37, ACM, 2003, pp. 31–47.
[5] N.J. Belkin, Some (what) grand challenges for information retrieval, in: ACM SIGIR Forum, 42, ACM, 2008, pp. 47–54.
[6] S. Calegari, G. Pasi, Ontology-based information behaviour to improve web search, Future Internet 2 (4) (2010) 533–558.
[7] L. Chen, K. Sycara, Webmate: a personal agent for browsing and searching, in: Proceedings of the Second International Conference on Autonomous Agents, ACM, 1998, pp. 132–139.
[8] Y. Chiaramella, Information retrieval and structured documents, in: Lectures on Information Retrieval, Springer, 2001, pp. 286–309.
[9] C. Chute, A. Manfrediz, S. Minton, D. Reinsel, W. Schlichting, A. Toncheva, The diverse and exploding digital universe, IDC White Paper, 2008.
[10] M. Daoud, L. Tamine-Lechani, M. Boughanem, Learning user interests for a session-based personalized search, in: Proceedings of the Second International Symposium on Information Interaction in Context, ACM, 2008, pp. 57–64.
[11] L.M. de Campos, J.M. Fernández-Luna, J.F. Huete, C.J. Martin-Dancausa, A. Tagua-Jiménez, C. Tur-Vigil, An integrated system for managing the andalusian parliament's digital library, Program: Electron. Lib. Inf. Syst. 43 (2) (2009) 156–174.
[12] L.M. De Campos, J.M. Fernández-Luna, J.F. Huete, A.E. Romero, Garnata: an information retrieval system for structured documents based on probabilistic graphical models, in: Proceedings of the Eleventh International Conference of Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU), 2006, pp. 1024–1031.
[13] L.M. de Campos, J.M. Fernández-Luna, J.F. Huete, E. Vicente-López, Using personalization to improve xml retrieval, IEEE Trans. Knowl. Data Eng. (TKDE) 26 (5) (2014) 1280–1292.
[14] S. Gauch, M. Speretta, A. Chandramouli, A. Micarelli, User profiles for personalized information access, in: The adaptive web, Springer, 2007, pp. 54–89.
[15] G. Gentili, A. Micarelli, F. Sciarrone, Infoweb: an adaptive information filtering system for the cultural heritage domain, Appl. Artif. Intell. 17 (8–9) (2003) 715–744.
[16] M.R. Ghorab, S. Lawless, A. O'Connor, V. Wade, Does personalization benefit everyone in the same way? Multilingual search personalization for english vs. non-english users., UMAP Workshops, 2014.
[17] K. Järvelin, J. Kekäläinen, Cumulated gain-based evaluation of ir techniques, ACM Trans. Inf. Syst. (TOIS) 20 (4) (2002) 422–446.
[18] D. Jiang, K.W.-T. Leung, L. Yang, W. Ng, Query suggestion with diversification and personalization, Knowl. Based Syst. 89 (2015) 553–568.
[19] A. Kobsa, Privacy-enhanced web personalization, in: The adaptive web, Springer, 2007, pp. 628–670.
[20] E.L.-C. Law, T. Klobučar, M. Pipan, User Effect in Evaluating Personalized Information Retrieval Systems, Springer, 2006.
[21] A.P. López-Monroy, M. Montes-y Gómez, H.J. Escalante, L. Villaseñor-Pineda, E. Stamatatos, Discriminative subprofile-specific representations for author profiling in social media, Knowl. Based Syst. 89 (2015) 134–147.
[22] M. Maleszka, B. Mianowska, N.T. Nguyen, A method for collaborative recommendation using knowledge integration tools and hierarchical structure of user profiles, Knowl. Based Syst. 47 (2013) 1–13.
[23] A. Micarelli, F. Sciarrone, Anatomy and empirical evaluation of an adaptive web-based information filtering system, User Model. User-Adapt. Interaction 14 (2–3) (2004) 159–200.
[24] A. Moukas, Amalthaea information discovery and filtering using a multiagent evolving ecosystem, Appl. Artif. Intell. 11 (5) (1997) 437–457.
[25] G. Pasi, Issues in personalizing information retrieval, IEEE Intell. Inf. Bull. 11 (1) (2010) 3–7.
[26] G. Salton, C. Buckley, Term-weighting approaches in automatic text retrieval, Inf. Process. Manage. 24 (5) (1988) 513–523.
[27] X. Shen, B. Tan, C. Zhai, Implicit user modeling for personalized search, in: Proceedings of the 14th ACM International Conference on Information and Knowledge Management, ACM, 2005, pp. 824–831.
[28] L. Shou, H. Bai, K. Chen, G. Chen, Supporting privacy protection in personalized web search, IEEE Trans. Knowl. Data Eng. (TKDE) 26 (2) (2012) 453–467.
[29] A. Sieg, B. Mobasher, R. Burke, Web search personalization with ontological user profiles, in: Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management, ACM, 2007, pp. 525–534.
[30] A. Spink, S. Ozmutlu, H.C. Ozmutlu, B.J. Jansen, Us versus european web searching trends, ACM Sigir Forum 36 (2) (2002) 32–38.
[31] B. Steichen, H. Ashman, V. Wade, A comparative survey of personalised information retrieval and adaptive hypermedia techniques, Inf. Process. Manage. 48 (4) (2012) 698–724.
[32] K. Sugiyama, K. Hatano, M. Yoshikawa, Adaptive web search based on user profile constructed without any effort from users, in: Proceedings of the 13th International Conference on World Wide Web, ACM, 2004, pp. 675–684.
[33] L. Tamine-Lechani, M. Boughanem, M. Daoud, Evaluation of contextual information retrieval effectiveness: overview of issues and research, Knowl. Inf. Syst. 24 (1) (2010) 1–34.
[34] X. Tao, Y. Li, N. Zhong, A personalized ontology model for web information gathering, Knowl. Data Eng. IEEE Trans. 23 (4) (2011) 496–511.
[35] J. Teevan, S.T. Dumais, E. Horvitz, Personalizing search via automated analysis of interests and activities, in: Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, 2005, pp. 449–456.
[36] J. Teevan, S.T. Dumais, E. Horvitz, Potential for personalization, ACM Trans. Comput. Human Interact. (TOCHI) 17 (1) (2010) 4.
[37] J. Trajkova, S. Gauch, Improving ontology-based user profiles., in: RIAO/OAIR: Recherche d'Information Assistée par Ordinateur / Open Research Areas in Information Retrieval, 2004, pp. 380–390.
[38] A. Trotman, B. Sigurbjörnsson, Narrowed Extended XPath I (NEXI), in: Advances in XML Information Retrieval, Springer, 2005, pp. 16–40.
[39] D. Vallet, P. Castells, M. Fernández, P. Mylonas, Y. Avrithis, Personalized content retrieval in context using ontological knowledge, IEEE Trans. Circuits Syst. Video Technol. 17 (3) (2007) 336–346.
[40] E. Vicente-López, L.M. de Campos, J.M. Fernández-Luna, J.F. Huete, A. Tagua-Jiménez, C. Tur-Vigil, An automatic methodology to evaluate personalized information retrieval systems, User Model. User-Adapt. Interact. 25 (1) (2015) 1–37.
[41] D.H. Widyantoro, J. Yin, M. El Nasr, L. Yang, A. Zacchi, J. Yen, Alipes: a swift messenger in cyberspace, in: Proceedings of Spring Symposium Workshop on Intelligent Agents in Cyberspace, 1999, pp. 62–67.
[42] Y. Yang, B. Padmanabhan, Evaluation of online personalization systems: a survey of evaluation schemes and a knowledge-based approach, J. Electron. Commerce Res. 6 (2) (2005) 112–122.
[43] L. Zighelnic, O. Kurland, Query-drift prevention for robust query expansion, in: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, 2008, pp. 825–826.