

Learning Parliamentary Profiles for Recommendation Tasks

Luis M. de Campos¹(✉), Juan M. Fernández-Luna¹, Juan F. Huete¹,
Pável Calado², and Bruno Martins²

¹ Departamento de Ciencias de la Computación e Inteligencia Artificial,
ETSI Informática y de Telecomunicación, CITIC-UGR, Universidad de Granada,
18071 Granada, Spain

{lci, jmfluna, jhg}@decsai.ugr.es

² Instituto Superior Técnico e INESC-ID, Universidade de Lisboa, Lisboa, Portugal
{pavel.calado, bruno.g.martins}@ist.utl.pt

Abstract. We consider the problem of building a content-based recommender system in a parliamentary context, which may be used for two different but related tasks. First, we consider a filtering task where, given a new document to be recommended, the system can decide those Members of the Parliament who should receive it. Second, we also consider a recommendation task where, given a request from a citizen, the system should present information on those deputies that are more involved in the topics of the request. To build the system we collected, for each Member of the Parliament, the text of corresponding speeches within the parliament debates and generated, with different techniques, a profile that was used to match against the input (document or request). We tested our methods using the documents of the regional Andalusian Parliament at Spain, obtaining promising results.

Keywords: User profiles · Content-based recommender systems · Information filtering · Information retrieval · Parliamentary documents

1 Introduction

The application of Information and Communication Technologies (ICT) in public administration, called e-government, aims to improve public services and democratic processes, as well as to strengthen support to the public [8]. Within this general context, we focus on the specific case of national or regional parliaments which, as other political institutions, produce a plethora of information but may fail to reach the people who are interested in it. This communication gap is related to citizens but also to the Members of the Parliament (MP) themselves [2]. As stated in [10] in the context of the European Parliament, “MPs need to be selective in their information input”. The same authors also stress the need to “increase dissemination of information and to provide opportunities for greater interaction between represented and representative”. Our proposals in this paper go in the direction of these two goals.

Let us assume that a new document must be distributed among the MPs. This document may be external to the parliament (e.g., a news release) or internal, such as a parliamentary initiative¹. The proposed system should be able to evaluate and decide those MPs who should receive it (a filtering task), according to their specific interests and the content of the document itself. On the other hand, consider an entity (e.g. a citizen or an institution) that wants to find the most appropriate MPs that can deal with a given request. In this case, the system should recommend who are these MPs, again taking into account their interests and the topical content of the request. In both cases, given an input to the system (either a new document that must be filtered or a request of a citizen), an automated procedure should determine those MPs who should receive this input.

Thus, our research may be framed within the field of content-based recommender/filtering systems [1,9], which are systems that recommend an item to a user, based on a description of the item and a profile of the user's interests. We are not aware of any previous description of one such system in a parliamentary context.

The source of information to learn about the interests of the MPs, in order to build the corresponding MP profiles, will be the transcriptions of all their speeches within parliament debates (either in plenary sessions or in more specific committee sessions). These transcriptions are collected in the form of records of proceedings.

The rest of the paper is organized in the following way: in Sect. 2 we describe the different ways we are going to use to build the MP profiles, as well as how these profiles will be used by an Information Retrieval System (IRS) to match them against the inputs, providing as a result a ranked list of MPs. The profiles will act as documents in a document collection and the inputs (documents or requests) will play the role of queries. In Sect. 3 we give details of the experimental framework designed to test our proposals, using a collection of records of proceedings of the regional Parliament of Andalusia at Spain. Section 4 ends the paper with the conclusions and some proposals for future work.

2 Building and Using Profiles

Considering the revision of the state-of-the-art about profiles and personalized IR given in [4], we are going to contextualize the type of profiles that we are working with in this research. The profiles that we present are built implicitly, mining relevant documents for the users, in our case, MPs who participate in parliamentary debates (in opposition to those constructed by explicit information given by users). With respect to the temporal perspective, our proposed profiles are static in the sense that they are not updated with the time (the alternative is the dynamic profile, which is changed when the preferences of the users change). In order to incorporate new information, our profiles have to be

¹ An initiative is the literal transcription of the discussion in the parliament of a petition presented by specific MPs or groups.

re-built from the scratch and we could say that they represent a long-term snapshot. The profiles at hand are only composed of a flat group of terms. We do not consider other type of information such as semantic networks or hierarchies, weighted concepts or association rules, for example.

Also in [4], the authors propose the following stages in the process of building profiles: (1) Collecting the user information; (2) Representing the profiles, and (3) Profile construction. Therefore, considering these phases, we are going to describe the decisions made to build the profiles, although previously we are going to explain how these profiles will be used after they are built.

2.1 Exploitation of the Profiles

It is important to remark that this paper, as stated in Sect. 1, is focused on the use of profiles for filtering tasks (and not for search personalization through query expansion [5]). In this case, we could consider that a citizen is looking for an MP who is involved in certain matters in which the citizen is interested in, expressed by means of a query to a retrieval system; or the case where given a new initiative that arrives to the system, we would like to know the names of those MPs who could be interested in it. In both cases, we have to use a more pure IR perspective, as the profiles would be considered documents from a classic IR view. The query could be the citizen's information need expressed as a set of keywords or the complete text included in an initiative. Then, an IRS indexing the document collection composed of the profiles would process the query, obtaining a ranked list of the MPs who are more involved with the content of the query as the result.

2.2 Collecting User Information

In this first stage, with the aim of extracting the information to build the profile, different mechanisms can be applied, as discussed in [4]. Taking into account that the context of this research are the parliamentary debates, and more specifically, their transcriptions of the MPs' interventions in the form of records of proceedings, as previously described in Sect. 1, we assume that we may build the MPs' profiles analyzing all their participations (speeches) in different initiatives discussed in committee or plenary sessions in the whole set of records of proceedings. For example, if an MP usually participates in debates related to agriculture, all their interventions will be around this topic and we could learn a profile composed of agriculture-related terms. That would be the case where the profile contains only one facet, but it is very easy to find the case where a politician is included in several commissions dealing with different topics, let us say education, environment and economics, for example. In these cases, the profiles originating from speeches could be considered multi-faceted and reflecting multiple preferences, as the MP is closer to several specific political areas.

Our approach considers, as the base for building the MPs' profiles, a collection of *documents*, each containing the whole text of all the speeches of a specific MP in the discussion of a particular initiative. We can represent as D the set

of documents, each compiling all the interventions of the i -th MP, i.e. $D = \{d_1, \dots, d_m\}$ (m being the number of MPs).

2.3 Representation

The second step is to determine the content of the profile and its internal organization. In our case, a specific MP's profile will be composed of a set terms extracted from the MPs' collection of interventions and finally weighted in order to reflect their importance or relevance in the complete profile. We assume term independence, so no type of relationships between terms is represented in the profile. Then, the profile of the i -th MP, P_i , is conceptually represented by a set of pairs term, weight, i.e. $P_i = \{(t_{i1}, w_{i1}), \dots, (t_{in}, w_{in})\}$, where n is the number of terms included in the profile. An important question that we shall partially try to answer in the experimentation phase is what is the most appropriate size of the profile with the aim of best representing the MP's ideas expressed in their speeches.

Noticing that not only the size is important, another research question that we tackle in this article is whether the composition of the profile in terms of parts of speech could be relevant for its quality. We shall take into account profiles composed of (1) only nouns (proper and common), which tries to simulate, in a certain way, concepts included in the interventions; (2) Nouns plus verbs, complementing the concepts with actions related to them; and finally, (3) all parts of speech (APoS). In this last case, we assume that stop words are not good language elements to be considered in a profile. As such, they are removed given a list of the most common words in the corresponding language (prepositions, articles, conjunctions, etc.), adding also a set of very frequent words in the regional parliamentary domain (for example, "deputy" or "Andalusia"). At this point, APoS profiles could contain nouns, verbs, some adjectives and adverbs, and also dates and numbers.

2.4 Construction

Broadly speaking, the process of constructing a profile P_i is basically founded on computing a weight for each term in the set of documents D , and selecting those n terms with the highest value. Then, profile P_i is built filtering terms from d_i . This would be a term selection task. Next, we have to focus our attention on the weighting schemes used to measure the importance of the terms in the profile, and to rank them accordingly. Two of the more well-known approaches are *tf* and *tfidf*. The former simply reflects the raw occurrence of the term in the collection of speeches, d_i , i.e. it means that the most frequent terms appearing in the transcriptions of each MP are selected. Profiles weighted with this scheme will just contain the most common words of each MP. The latter, the *tfidf*, takes into account, not only term frequencies in documents but also their rarity within the whole collection. The *idf* component of a term t is the inverse document frequency computed as $\log\left(\frac{|D|}{df_t}\right)$, where df_t is the number of

documents where term t occurs. Profiles, which terms are selected based on this scheme, will contain a vocabulary more particular of the MP, weakening terms which appears very frequently in the speeches of other MPs.

In addition to these approaches, we add a new one, called *Difference* ($Diff$), presented in [11] in the context of profiles used to personalize search results via query expansion. The foundation of this new weighting scheme is basically that if a term t is more frequent inside a document d_i than outside (the rest of documents), the term is representative of that i -th MP. Otherwise, if the term occurs more times in the speeches of other MPs than in the i^{th} MP, then it is not useful for her as a representative of her speeches. More formally, we define $f^+(t, d_i)$ as the frequency of t in d_i ; $f^+(d_i)$ is the total number of terms in d_i ; $f^-(t, d_i)$ and $f^-(d_i)$ are respectively the frequency of t and the number of terms in documents outside d_i . We then define the difference measure of t with respect to d_i , $Diff(t, d_i)$, as:

$$Diff(t, d_i) = \frac{f^+(t, d_i)}{f^+(d_i)} - \frac{f^-(t, d_i)}{f^-(d_i)} \quad (1)$$

The previous equation corresponds to the normalized frequency of t within d_i minus the normalized frequency of t outside d_i . If the final value is $Diff(t, d_i) \leq 0$, it means that t is more frequent outside than within d_i , so it is not representative of d_i . However, if the final value is $Diff(t, d_i) > 0$, this means that t represents d_i at a certain degree.

Using any of these three approaches we obtain as the output, for each MP, a set of n selected terms together with their corresponding weights.

Finally, and in order to facilitate the use of the profiles for any given search engine, we have decided not to use directly the sets of selected weighted terms as the profiles. Instead, we build each profile as a bag of words, replicating the selected terms several times, simulating the way a document contains term occurrences. This facilitates the profiles to be indexed by existing search engines, and therefore we can use their internal weighting schemes.

We propose two different alternatives for this purpose: The first one, noted as *R- tf* , is to replicate any selected term as many times as its tf (number of occurrences in d_i), reflecting its real occurrence pattern, independently on its weight. The second one, *R-Prop*, is designed according to the fact that if a term has got a higher weight, then it should be replicated more times than those with lower weights. To carry out this idea we apply a linear transformation to the original weight, considering the maximum and minimum weight values as well as the number n of terms to be included in the profile. This transformation is: $replications = (int) \left((n - 1) \frac{weight - min_weight}{max_weight - min_weight} + 1 \right)$. Therefore, the term having maximum weight will have a replication value of n , while the one with minimum weight will have a replication value of 1. This way is intended to give occurrences of a term proportional to its weight.

3 Evaluation

In this section we describe the components of the evaluation framework, as well as the obtained results and conclusions. In order to evaluate the quality of the different profiles we shall try to measure their capability at predicting whether an MP might have some interest in a new initiative (considering only the content of the speeches). So, given one such initiative, the system will give as output a ranking of those MPs having the profiles that best match the target initiative (particularly, we shall focus on the top-10 MPs).

3.1 Evaluation Framework

The evaluation framework is composed by the following components:

Data Set: We use all the initiatives (5258) from the 8th term of office of the Andalusian Parliament at Spain², marked up in XML [3]. These initiatives contain a set of 12633 different interventions. From this set we shall build the profiles only for those deputies (or technical guests) who participate in more than 10 different initiatives, giving a total of 132 profiles.

Parameter Configuration: The different parameters that will be considered in these experiments are³: the number of terms in the profile, which takes values 50, 250, 500, 750 and 1000; The Part of Speech (the type of terms in the profile), which might be Noun (N), Noun+Verb (NV) and all PoS (A); The metrics for selecting the best terms in the profile are *tf* (t), *tfidf* (i) and *Diff* (d), and finally, the importance of the term in the profile that will be either *R-tf* (F) or *R-prop* (P). In total, a set of 90 different types of profile were built (plus one that consists of using all the terms in the MPs' speeches). We use the profiles in conjunction with the search engine library Lucene⁴, considering three different similarity measures, namely BM25, Language Model and Cosine, giving a total number of 273 different evaluations.

Evaluation Methodology: In order to evaluate the performance of our proposals, we have used the repeated holdout method [7]: the set of initiatives is randomly partitioned into training (80%) and test (20%), and we repeat this process five times (the results presented in the study are the averages over the different rounds). The training set is used to learn the MP profiles. With respect to the initiatives in the test set, we remove the information related to the MPs who participate in their discussion, being totally anonymous. In these experiments we have considered, as ground truth, that a test initiative, j , will be relevant only to those n_j MPs who participate in it⁵.

² <http://www.parlamentodeandalucia.es>.

³ In parentheses we present the short acronym used in the figures and tables in this section.

⁴ <https://lucene.apache.org>.

⁵ We consider that this is a rather conservative assumption, because it is quite reasonable to think that an initiative can also be relevant to other MPs.

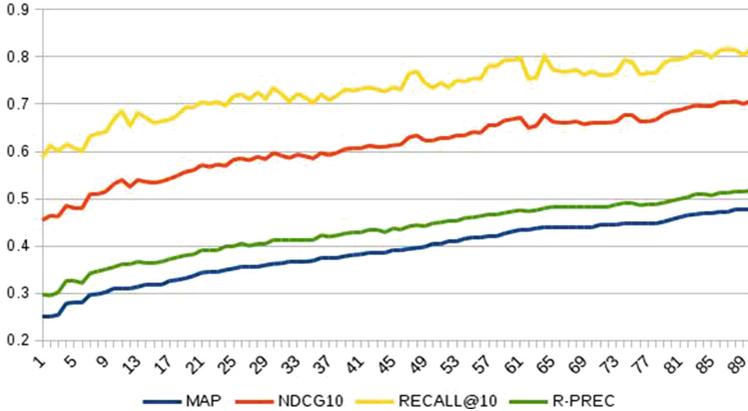


Fig. 1. Performance for the different evaluation metrics using BM25.

Evaluation Metrics: In order to evaluate the accuracy of our approach, we use the recall values considering the top-10 MPs, $\text{rec}@10$, that measures how many MPs, among the relevant MPs, appear in the top positions of the ranking and the Normalized Discounted Cumulative Gain [6], $\text{NDCG}@10$, which measures the ranking quality. Also, considering that the number of relevant MPs varies with the initiative (on the average there are 2.4 interventions per initiative) we shall also measure the accuracy on the top n_j positions using $\text{MAP}@n_j$ (Mean Average Precision) and the R-precision (that represents the precision over the top n_j MPs).

3.2 Results

Although the selected metrics measure different features in the output ranking, it deserves to be mentioned that they all are highly correlated, with correlation values greater than 0.97 among all of them, independently of the used similarity function. In order to illustrate this fact, Fig. 1 presents the values obtained by each metric for each one of the 90 possible parameter configurations (in the x-axis) using BM25. We believe that this fact strengthens the conclusions obtained in this study because the results are consistent among the different metrics. Also, this helps us to reduce the number of results to be shown, focusing on the results obtained with one of them; particularly we have selected the $\text{NDCG}@10$ as the main metric through the main part of this section. We shall organize this section through 3 different research questions, the first two related to the parameter configuration and the last one devoted to analyzing the differences between the similarity measures.

Q1: Which is the right number of terms of the profile? In order to answer this question, we have run an intensive test considering different profile sizes, fixing the rest of the parameters to the configuration which gives the best results

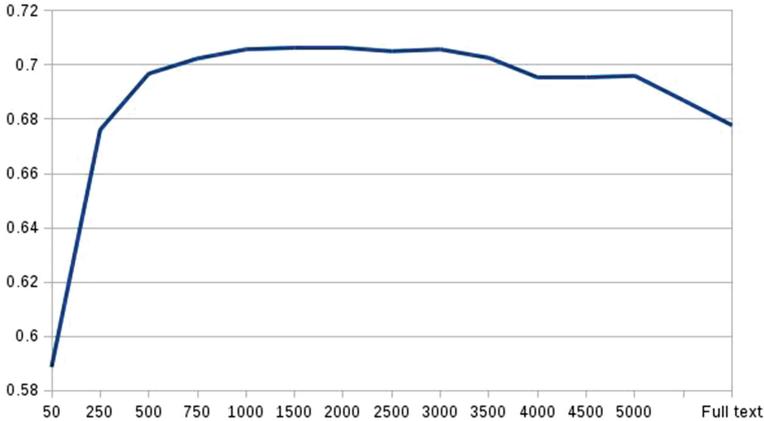


Fig. 2. NDCG@10 values, using BM25, when varying the size of the profile.

(discussed below). The results for the NDCG@10 metric and BM25 are presented in Fig. 2. From this figure we can see that using a few number of terms the results are not good enough, increasing to the best values using a profile size in the range 1000 to 5000 terms (which is better than considering all the terms), then the results decrease when the profile size increases.

From this result we can conclude that the best alternative is to consider a profile with around 1000 terms, being a good balance between size and quality of the results. Also, our experiment shows that for our proposal it is necessary to consider a huge number of terms in the profile (the results obtained using only 50 terms are quite poor). Note that this fact is against the profile's size necessary for helping the users to improve the search experience, i.e. to find those documents relevant to a given query. In this case, the use of large profiles is not convenient because of the query drift problem⁶. This also suggests that the search of the best profile in a filtering task is a different problem than, for instance, successfully personalize search results.

Q2: Which is the best configuration of profile parameters? In order to answer this question, we shall use the net chart presented in Fig. 3. This chart shows the NDGC@10 values for each candidate configuration when using BM25 as similarity measure. In the grid all the different accuracy values obtained with the same profile size, i.e., 50, 250, 500, 750 and 1000 have been connected by lines, so we can see easily how the performance varies with the configuration.

From this chart we can say that the best configuration is:

size = 1000; type = APoS, selection = *tfidf*, weight = *R-prop*.

⁶ The change in the underlying *intent* between the original query and its expanded form.

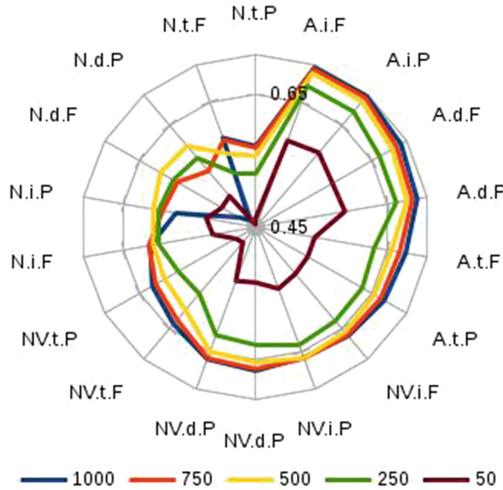
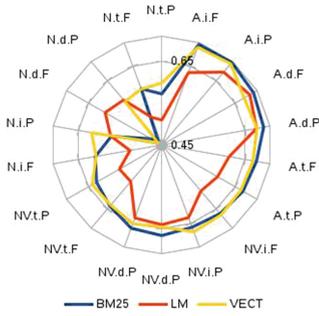


Fig. 3. NDCG@10 values for the profile sizes equals to 50, 250, 500, 750 and 1000 using BM25 as similarity metric.

Although this is important, other conclusions can also be obtained: Firstly, with respect to the part of the speech considered for selecting the terms in the profile, it seems that it is better not to consider any restriction about the selected terms, i.e. all the parts of the speech matter. Nevertheless, the tendencies in the results are different when using only nouns⁷ versus the use of nouns and verbs or all PoS. Secondly, the best alternative for selecting terms for being included in the profile is to take into account *tfidf* weights, although there are not significant differences with respect to the use of Diff. In this sense, the results get worse when only the frequencies are taken into account. In other words, how helpful is a term to describe the user preferences does not depend only on the raw frequency of the term in the speech, but it also depends on how common is this term in the rest of the speeches, fact that has been taken into account by the other two metrics. It is worth to remark that this holds for all types of terms. Thirdly, there is not a clear winner with respect to the use of a frequency or proportional-based approaches to determine the weight of the terms in the profile (there is not statistical significance between the results, using a two-tail t-test we obtain a p-value equals to 0.00193). In this case, the best results have been obtained using a proportional criterion, although this criterion only wins in the 37.7% of the configurations.

Q3: Is there any difference between the different similarity models? In order to tackle this question we shall discuss the results obtained using BM25, Language Model (LM) and Cosine (VECT), fixing the profile size to 1000. Figure 4

⁷ We are not going to discuss their impact in this work, since the accuracy values obtained are worse in all the cases.



Param.	rec@10	NDCG	MAP	R-prec
BM25				
Full	0.7770	0.6778	0.4568	0.4959
A.i.P	0.8177	0.7074	0.4773	0.4876
%	5.23	4.37	4.48	-1.63
LM				
Full	0.7421	0.6176	0.3867	0.4329
A.d.F	0.8095	0.6903	0.4542	0.4953
%	9.08	11.77	17.45	14.41
VECT				
Full	0.7971	0.6876	0.4584	0.5030
A.i.P	0.8231	0.7036	0.4658	0.5087
%	4.39	2.32	1.61	1.13

Fig. 4. Results for the different similarity metrics, i.e. BM25, LM and VECT.

shows a net chart with the obtained NDCG@10 values and also a table with all the values obtained by the best parameter configuration. In this table we also present the results when we do not learn a profile, i.e. all the terms in the speech are considered (denoted as Full) that can be considered the baseline, and the improvements (in terms of percentage of gain, %) for each metric.

In this case, we can see that both BM25 and VECT exhibit a similar performance (particularly, when considering APoS or N+V) whereas LM behaves differently in tendencies and results. Thus, in the case of LM it seems that it is better to consider *Diff* as criterion guiding the term selection process (this is in some way congruent with the criterion used by LM to compute the terms weights). Also, it is interesting to note that although LM gives the worst results, greater improvements (with respect to the baseline) are obtained.

4 Conclusions

We have designed a new alternative to build user profiles in a Parliamentary context, which is based on considering parliamentary interventions as bags of words. In this sense, the profiles can be considered as new documents that can be indexed by a search engine. In this paper we have explored the effects of three features on the profile quality: the number of terms, their grammatical type and their associated weights (encoded as the frequency of a term in the profile). Our experimental results show that it is better to consider a large number of terms for describing the users’ preferences, which suggests that a profile for filtering has different features than the ones used in search tasks. We have also explored the effect of the parts of the speech used in the profile, showing a strong evidence in favor of the use of all parts of the speech. Also, our results show that, in order to select the best terms, it is necessary to consider how common is a term in the other MPs speeches.

Although our current focus is on a parliamentary context, the ideas developed here could be used in other situations, for instance in recommending new technical papers to scientists, or in discovering those scientists who are more

prominent for a given topic. For future work, we plan to tackle the fact that the profiles can be heterogeneous and also the impact of the time in the profile, i.e. analyzing short-term vs long-term users' interests.

Acknowledgements. Paper supported by the Spanish “Ministerio de Economía y Competitividad” under the project TIN2013-42741-P.

References

1. Belkin, N.J., Croft, W.B.: Information filtering and information retrieval: two sides of the same coin? *Commun. ACM* **35**, 29–38 (1992)
2. Busby, A., Belkacem, K.: ‘Coping with the Information Overload’: An Exploration of Assistants’ Backstage Role in the Everyday Practice of European Parliament Politics. *European Integration online Papers*, vol. 17 (2013)
3. de Campos, L.M., Fernández-Luna, J.M., Huete, J.F., Martin-Dancausa, C.J., Tur-Vigil, C., Tagua, A.: An integrated system for managing the andalusian parliament’s digital library. *Program Electron. Libr. Inf. Syst.* **43**, 121–139 (2009)
4. Gauch, S., Speretta, M., Chandramouli, A., Micarelli, A.: User profiles for personalized information access. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) *Adaptive Web 2007*. LNCS, vol. 4321, pp. 54–89. Springer, Heidelberg (2007)
5. Pasi, G.: Issues in personalizing information retrieval. *IEEE Intell. Inf. Bull.* **10**, 3–7 (2010)
6. Jarvelin, K., Kekalainen, J.: Cumulative gain-based evaluation of IR techniques. *ACM Trans. Inf. Syst.* **20**, 422–446 (2002)
7. Lantz, B.: *Machine Learning with R*. Packt Publishing Ltd., UK (2013)
8. Palvia, S.C.J., Sharma, S.S.: E-government and e-governance: definitions/domain framework and status around the world wide web. *Foundations of e-government*. In: *5th International Conference on E-Governance*, pp. 1–12 (2007)
9. Pazzani, M.J., Billsus, D.: Content-based recommendation systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) *Adaptive Web 2007*. LNCS, vol. 4321, pp. 325–341. Springer, Heidelberg (2007)
10. Shamin, J., Neuhold, C.: ‘Connecting Europe’: the use of ‘new’ information and communication technologies within european parliament standing committees. *J. Legislative Stud.* **13**, 388–402 (2007)
11. Vicente-López, E., de Campos, L.M., Fernández-Luna, J.M., Huete, J.F.: Personalization of parliamentary document retrieval using different user profiles. In: *Proceedings of the 2nd International Workshop on Personalization in eGovernment Services and Applications (PEGOV2014)*, pp. 28–37 (2014)