Generic knowledge-based analysis of social media for recommendations

Victor de Graaff
Dept. of Computer Science
University of Twente
Enschede, The Netherlands
v.degraaff@utwente.nl

Anne van de Venis
Dept. of Computer Science
University of Twente
Enschede, The Netherlands
a.j.vandevenis@student.utwente.nl

Maurice van Keulen
Dept. of Computer Science
University of Twente
Enschede, The Netherlands
m.vankeulen@utwente.nl

Rolf A. de By
Fac. of Geo-Information Science
& Earth Observation (ITC)
University of Twente
Enschede, The Netherlands
r.a.deby@utwente.nl

ABSTRACT

Recommender systems have been around for decades to help people find the best matching item in a pre-defined item set. Knowledge-based recommender systems are used to match users based on information that links the two, but they often focus on a single, *specific* application, such as movies to watch or music to listen to. In this paper, we present our *Interest-Based Recommender System* (IBRS). This knowledge-based recommender system provides recommendations that are *generic* in three dimensions: IBRS is (1) domain-independent, (2) language-independent, and (3) independent of the used social medium. To match user interests with items, the first are derived from the user's social media profile, enriched with a deeper semantic embedding obtained from the generic knowledge base DBpedia. These interests are used to extract personalized recommendations from a tagged item set from any domain, in any language. We also present the results of a validation of IBRS by a test user group of 44 people using two item sets from separate domains: greeting cards and holiday homes.

Keywords

Recommender systems, knowledge-based, DBpedia, social media, domain-independent, language-independent

General Terms

Algorithms, Design, Experimentation

Categories and Subject Descriptors

H.4.2 [**Information Systems Applications**]: Types of Systems—*Decision support*

1. INTRODUCTION

The aim of a *recommender system* (RS) is to help people find the items they are most interested in. A requirement to provide *personalized* recommendations is that the RS has knowledge of the *person* using it. In 2013, Facebook claimed to have 1.11 billion active users [1], and the top-100 *pages* alone currently have a total of 5.87 billion *facebook-likes* [2]. The items that people express a preference for on social media, whether through a *like* of a Facebook page, a *follow* on Twitter, or a *tip* on the renewed FourSquare, can be taken to disclose personal traits of interest and the things they want to be associated with. This vast amount of information is the starting point for our *Interest-Based Recommender System* (IBRS).

But what people express their preference for on social media, cannot always directly be related to commonly used tags or words in descriptions in an existing item set. These items are often example instances of broader concepts. For example: *Cristiano Ronaldo* has 103 million *facebook-likes* at the time of writing, whereas Soccer (66 million) and Football (46 million) have considerably fewer facebook-likes.[1] Tag sets or descriptions, on the other hand, are more likely to contain these broader concepts, as for example is the case in greeting cards, sports equipment, or campsites with soccer fields. In fact, one of our validation item sets contains tagged greeting cards with practically only generic terms such as soccer/football. To bridge this generalization gap in a domain- and language-independent way, we use the multilingual, generic knowledge base DBpedia to automatically detect broader concepts. We call these concepts the user's *interests*. In this paper, we validate our hypothesis that automated user interest detection can also be used to select preferred items in an item set, independent of the item set domain, language and used social medium. As a boundary requirement to our solution, the cold-start problem, as for example discussed by Bobadilla et al. [3], needs to be circumvented. The system we propose shall be seen as a feature of a larger recommender system, either to bootstrap or to support that system, rather than as a stand-alone system.

In addition to the recommendation approach we propose in this paper, we also present the results of a validation thereof. A user group of 44 people tested our RS, using item sets from two completely different domains: greeting cards and

CBRecSys 2015, September 20, 2015, Vienna, Austria.

[1] Synonyms like this one cause problems as well, and are discussed in more detail in Section 3

holiday homes. Both the recommendation selection, as well as the explanation interface were validated by these users, using their own social media profile.

This paper is further structured as follows: related work is discussed in Section 2, the motivation behind this research is discussed in Section 3, the IBRS technology is presented in Section 4, while the validation approach and results are laid out in Section 5, and Section 6 finally contains concluding remarks and hints at future work.

2. RELATED WORK

The creation of a RS that makes use of social media or DBpedia is not a new ambition. Social media have especially received much attention in the field of content-based recommender systems. Fijałkowski and Zatoka presented an architecture of a recommender system for e-commerce based on Facebook profiles [4]. Guy et al. proposed five recommender types, based on social media and/or tags [5]. In their approach, they also presented the users with recommendation explanation. The social media they focus on however, are not of the mainstream type, but specific for the Lotus Connections suite. The system of He et al., on the other hand, uses common social media [6]. Whereas they claim to overcome the cold-start problem, their system appears to still suffer from the new item cold-start problem, as described by Bobadilla et al. [3].

The creation of a RS based on DBpedia has also received quite some attention already, especially in the field of music [7, 8] and movie [9, 10, 11, 12, 13] recommendation. Di Noia et al. took it a step further and also benefited from the integration of DBpedia in the *linked open data* (LOD) initiative. Their movie recommendations are not only based on DBpedia knowledge, but also on Freebase and LinkedMDB. A more generic approach to create a RS using LOD was done by Heitmann and Hayes [14], who use also use LOD to overcome the cold-start problem. Even though their validation is based on a music dataset, their approach has the genericity to be used for other applications as well. Our approach for broader concept detection through DBpedia is a form of knowledge-based query expansion. Liang et al. already showed in [15] that document recommendation based on the user's interests improves as a result of query expansion, or semantic-expansion as they call it.

What distinguishes our approach from other RS research, is that we use both social media profiles and DBpedia data to create a *generic* RS. Passant and Raimond, for example, created a RS based on exported social media profiles and DBpedia data in [8], but their approach is limited to the music-specific relations in DBpedia. To the best of our knowledge, the only other generic approach is TasteWeights by Bostandjiev et al. [16]. They build a user profile based on social media data, and then apply a collaborative filtering-based approach to select recommendations. This still implies all of the three cold-start problem categories: new item, new user, and new community, again as described by Bobadilla et al. [3]. As it is exactly our goal to overcome the cold-start problem, our approach is a hybrid between content-based and knowledge-based, according to the RS classification by Burke and Ramezani [17]. Basile, Lops et al. would classify our work as a top-down *semantics-aware* content-based RS [18, 19].

Our work is inspired by Shi et al.'s HeteRecom [20], which is based on the similarity calculation HeteSim [21]. Similar to their work, our ultimate goal is to find the matching paths between a user and the item set that carry the most weight. In this paper however, we focus on the detection of existing paths.

3. MOTIVATION

In this work, we aim to extract recommendations that are generic in three dimensions: the recommendation approach shall be independent of the item set domain, the item set language, and the used social medium. As a fourth criterium, it shall not suffer from any of Bobadilla's three cold-start problem categories. Below, we discuss the motivation for all of these challenges:

Domain-independence

As discussed in the previous section, currently most recommender systems based on knowledge bases and social media are focused on one specific domain. Independence of the item set domain only allows us to reuse the solution and its future improvements for multiple applications.

Language-independence

Similar to domain-independence as a requirement for reusability, a language-independent solution improves the RS's potential to be used in multiple applications. A sub-requirement of of language-independence is synonym-independence. As Zanardi and Capra pointed out in [22], synonyms are a typical RS problem, especially for tag-based RSs. The example of people facebook-liking either the *Soccer* page or the *Football* page from Section 1 already showed that people may facebook-like different pages, while referring to the same concept. Despite recent efforts by Facebook to merge pages about the same topic from different languages into one page, and improving the search functionality to help people finding such pages while searching for their name in a different language, still several pages exist to describe similar concepts.

Social medium-independence

From the first form of genericity, domain-independence, follows another requirement. Several social media, such as Facebook, LinkedIn, Twitter, Instagram, and Pinterest, are widely used, and each of these has its own focus. When one decides to create a RS for job vacancies, LinkedIn may be a more logical social medium to base the recommendations on than any of the other, while a RS for touristic hotspots will most likely lead to another choice. Therefore, to create a RS based on social media content that is domain-independent, it shall also be independent of the underlying social medium.

Cold-start problem

The cold-start problem has been widely discussed in RS literature. Bobadilla et al. categorized it into three subcategories: the new item problem, the new user problem, and the new community problem [3]. Knowledge-based RS have been designed to overcome all of these problems, but often require domain-specific knowledge.

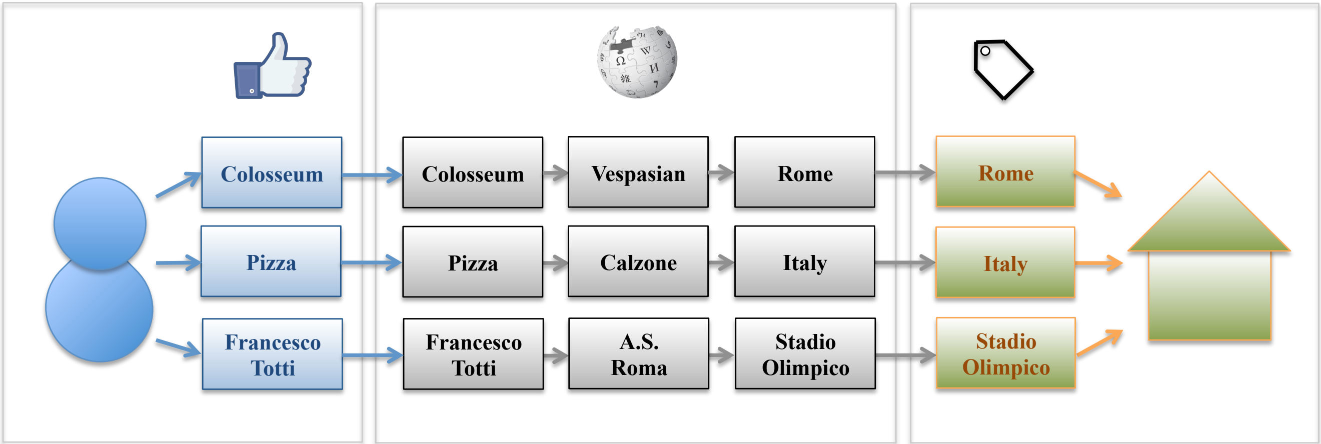


Figure 1: The IBRS concept, illustrated using the holiday home domain. A user's preferred items on social media are mapped onto knowledge base resources. Broader concepts are detected by exploring the knowledge base graph, and finally mapped onto tags in the item set database.

Overcoming all of these four challenges at the same time has motivated us to create IBRS: a domain-independent, language-independent, social medium-independent, knowledge-based RS.

4. CONCEPT & TECHNOLOGY

The foundation of IBRS is the idea that people are more likely to be interested in items that have a not too distant relation with things we know they like. Although things people express a preference for on social media are typically in a different domain than our item set, they may still give hints towards a person's interests. In IBRS, we link the preferred items on social media to resources in the DBpedia *Resource Description Framework* (RDF) graph. We use this graph to explore related concepts, which are then matched with a known *tag set*, that is used to label the *item set*. As a final step, we rank the item set based on the number of *matched tags*. This concept is illustrated, using the holiday home domain, in Figure 1. In this example, the user facebook-liked the Colosseum, pizza, and Francesco Totti. These facebook-likes are mapped onto DBpedia, and the DBpedia RDF graph is explored to detect the broader concepts Rome, Italy, and Stadio Olimpico. These items are mapped onto holiday home tags, to ultimately match the user with a specific holiday home.

The remainder of this section is structured as follows: RDF graph exploration is discussed in Section 4.1. The data model of the IBRS abstraction layer is presented in Section 4.2. Section 4.3 presents a method for automated tag generation from descriptions. In Section 4.4 the ranking mechanism and Facebook-DBpedia mapping approach are presented. Section 4.5, finally, presents a short introduction of the IBRS prototype.

4.1 DBpedia graph exploration

After matching a facebook-like with a DBpedia resource, we traverse the RDF graph in exactly two steps. Since RDF tuples have a *subject*, *predicate* and *object*, RDF graphs are directed. Therefore, there are four possible different direction combinations to travel from node A through node B to its second neighbor C.[2] In Table 1, we show the top-10 of second neighbors when traversing the DBpedia graph starting from the *Eiffel Tower* as node A, using all four possible direction combinations. DBpedia pages in italics also occur as tags in at least one of our two validation sets, which are discussed in detail in Section 5. The first approach, $A \rightarrow B \rightarrow C$, leads to results describing France, influential French people, and several other buildings in France. The second approach, $A \leftarrow B \rightarrow C$, has some overlap with the first approach, but also contains several results unrelated to France, such as Los Angeles and the United States. The third approach, $A \leftarrow B \leftarrow C$, shows some remarkable buildings throughout Europe, but also very unrelated lists towards the bottom of the top-10. The fourth and final approach, $A \rightarrow B \leftarrow C$, results in several famous French people, especially scientists. Other starting points show similar results: the third approach, $A \leftarrow B \leftarrow C$, shows promising results for single domain recommendations, whereas the first approach shows the best results for broader concept detection. Since our aim is to match these second neighbors with a tag set, we use the first approach, $A \rightarrow B \rightarrow C$.

4.2 Abstraction layer data model

To ensure IBRS genericity, an abstraction layer is used on top of the underlying data source, such as a product database. This abstraction layer can consist of physical tables, views, or a mix thereof, but we will refer to its items as tables from here on. The abstraction layer contains two entity tables: ABSTRACT_ITEMS and TAGS, and one relationship table: ABSTRACT_ITEMS_TAGS, as depicted in Figure 2.

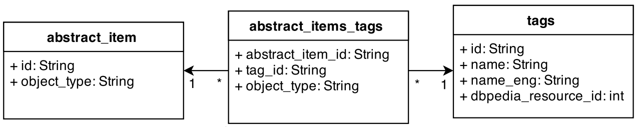


Figure 2: Abstraction layer data model

The ABSTRACT_ITEMS table contains the ID and OBJECT_TYPE

[2] Depending on the directions of the relationships, and the existence of bi-directional relationships, node A may be equal to node C, as can also be seen in Table 1.

Rank	$A \rightarrow B \rightarrow C$ (#)	$A \leftarrow B \rightarrow C$ (#)	$A \leftarrow B \leftarrow C$ (#)	$A \rightarrow B \leftarrow C$ (#)
1	*Paris* (20)	*Eiffel Tower* (41)	*Eiffel Tower* (7)	Paul Langevin (51)
2	*France* (20)	*France* (17)	Palácio de Ferro (3)	Léon Foucault (48)
3	*Eiffel Tower* (7)	*Paris* (15)	Cologne Cathedral (2)	Jean Témerson (48)
4	Manuel Valls (6)	*Los Angeles* (4)	Eiffel Bridge, Ungheni (2)	Frédéric Passy (45)
5	François Hollande (6)	British Library (4)	Souleuvre Viaduct (2)	L.A. de Bougainville* (45)
6	Unitary state (6)	Bonnétable (4)	Samuel Hibben (2)	Cecile de Brunhoff (45)
7	*French language* (6)	Aarhus University (4)	Casa de Fierro (2)	Adrien-Marie Legendre (45)
8	Anne Hidalgo (6)	Garabit viaduct (4)	Modern Marvels episodes* (2)	Robert Perrier (45)
9	Bonnétable (4)	St Paul's Cathedral (4)	Monopoly editions USA* (2)	Paul Lévy (math.)* (45)
10	Garabit viaduct (4)	*United States* (4)	Garabit viaduct (2)	Émile Drain (45)

Table 1: Top-10 of second neighbor nodes C through DBpedia graph exploration in multiple directions for the *Eiffel Tower* resource as node A. Numbers between brackets indicate number of paths between that node and the Eiffel Tower node. Items in italics also occur as tags in at least one of our two validation tag sets. Items marked with an asterisk are abbreviated.

of the items in the item set. The OBJECT_TYPE field allows us to use one IBRS instance for the recommendation of multiple item sets.

The TAGS table contains the tag's ID, NAME, and DBPEDIA_RESOURCE_ID. The NAME field can be used in the language of the item set tags. Since we have one item set that is tagged in Dutch, and one item set that is tagged in English, we added the NAME_ENG field for English tags. The DBPEDIA_RESOURCE_ID is cached in the database for better performance.

The ABSTRACT_ITEMS_TAGS table is a regular relation table containing the ABSTRACT_ITEM_ID and TAG_ID. It also contains the abstract item TYPE for improved join executions.

4.3 Tag generation

In case an item set is not tagged, but does contain descriptive texts, tags can be extracted automatically. Natural language processing algorithms can be used for this purpose, such as the named entity extraction and disambiguation approach by Habib et al. [23]. We used Habib's approach with a manually trained model to extract named entities from holiday home descriptions. A drawback of this approach is that descriptions are often the result of free-text input. Phrases such as "only a 3 hour flight from Amsterdam" or "25 kilometers from the border with France" led to correctly extracted named entities, but semantically not the best tags to distinguish this object from others. Therefore, we additionally removed those tags that tagged a holiday home with another country than the one it is located in. In total, this approach allowed us to assign 455,777 (non-unique) tags to 42,148 holiday homes, from which 106,430 tags (of which 12,151 unique) could be mapped onto a DBpedia resource.

4.4 Ranking

The IBRS ranking method consists of four steps: (1) retrieving preferred items from social media, (2) matching these items with DBpedia resources, (3) extracting abstracts from DBpedia, (4) ranking items based on matched tags. For performance reasons, several items are cached offline.

Obtaining preferred items from social media

To map social media items while remaining independent of the social medium, we must take into account that not all APIs are the same. Some social medium APIs allow developers to find out what a user's friends prefer, while others limit the developer to information about the logged in user. Therefore, when using the Facebook Graph API, we limited ourselves to the NAME and CATEGORY elements of each facebook-liked page.

Matching social media items with DBpedia resources

Facebook-likes are mapped onto DBpedia resources through their name. Those facebook-pages that mapped onto ambiguous terms in DBpedia were filtered out. To create a more complete mapping, we used the category element to postfix the name of those pages pages for which the category element was filled with "movie," "tv show," or "musician/band." In these cases, we also checked if a page exists with the additional suffix "_(movie)," "_(TV_series)," or "_(band)" respectively. This leads to the following SPARQL query:

```
PREFIX dbpont: <http://dbpedia.org/ontology/>
PREFIX dbpres: <http://dbpedia.org/resource/>
# We use the prefixed versions here for readability

SELECT ?uri ?label
WHERE {
 # Find exact match with category suffix
 { ?uri dbpont:wikiPageID [].
     FILTER(?uri = dbpres:The_Net_(movie)) }

 # Or exact match without category suffix
 UNION { ?uri dbpont:wikiPageID [].
     FILTER(?uri = dbpres:The_Net) }

  # Or the label version
 UNION {?uri rdfs:label "The_Net"@en.}

 # Check if page has redirect
 UNION { dbpres:The_Net_(movie)
     dbpont:wikiPageRedirects ?uri}
 UNION { dbpres:The_Net
     dbpont:wikiPageRedirects ?uri}

 ?uri rdfs:label ?label.
 ?uri dbpont:wikiPageID ?wikiPageid.
 FILTER (langMatches(lang(?label),"en")).

 # Filter out ambiguous terms
 FILTER NOT EXISTS { ?uri
   dbpont:wikiPageDisambiguates ?disambiguates } .
 # Filter out Wikipedia categories
 MINUS {?uri rdf:type skos:Concept}
}
LIMIT 1
```

Using this approach on a test set of 11,674 unique Facebook pages, obtained from the likes of 309 users, we were able to match 2,240 (19.2%) Facebook-pages with a DBpedia resource.

Extracting abstracts from DBpedia

For all matched DBpedia resources, the abstracts are retrieved from the SPARQL endpoint provided by DBpedia [24] using the following query:

```
PREFIX dbpont: <http://dbpedia.org/ontology/>
PREFIX dbpres: <http://dbpedia.org/resource/>

SELECT DISTINCT
 ?o3 (count(?o3) as ?count) ?abstract ?label

WHERE {
 # UNION concatenation of mapped FB pages
 {dbpres:Vienna ?p1 ?o2} UNION
 {dbpres:Recommender_system ?p1 ?o2} UNION
 {dbpres:Computer_science ?p1 ?o2}

 # Neighboring object has Wikipage
 ?o2 dbpont:wikiPageID ?o2id ;

 # Neighboring object has neighbor
    ?p2 ?o3 .

 # Second neighbor object has Wikipage
 ?o3 dbpont:wikiPageID ?o3id ;
    dbpont:abstract ?abstract ;
    rdfs:label ?label .

 # English is used as an example
 FILTER(langMatches(lang(?abstract), 'en')) .
 FILTER(langMatches(lang(?label), 'en')) .

 # Second neighbor object must not be a category
 MINUS {?o3 rdf:type skos:Concept}
}

# 'Only' the 1000 most important abstracts
ORDER BY DESC(?count)
LIMIT 1000
```

Ranking items based on matched tags

Each tag that (1) has a DBPEDIA_RESOURCE_ID and (2) is contained in at least one of the downloaded abstracts, is marked as a *matched tag*. The item set is then ranked on the basis of the number of matched tags. As a final step, those items that are too close to a higher ranked item, based on a pre-defined distance function, are removed from the ranking. This last step is added to ensure diversity among the recommended items. For the recommendation of geographic objects, as for example in a geo-social RS like the one discussed in [25], one can think of the Euclidean distance, but for more generic purposes the cosine similarity (as for example discussed in [22]) of the item's tags may be a good starting point. The tag input makes our RS domain-aware. However, since the approach can be applied to any tag domain, we still consider the concept itself domain-independent. This in contrast to for example music recommenders that rely on the artist-song relationship.

4.5 Prototype

For demonstration and validation purposes, we have created a prototype of IBRS, using the Cake PHP platform. The prototype can be used with either one's own Facebook profile, or by manually combining several DBpedia resources. It can be accessed through `http://ibrs.ewi.utwente.nl`.

5. VALIDATION

To validate our ranking mechanism, as well as to determine the user perception of recommendations with explanations, we validated IBRS in a carefully designed user study with a test user group of 44 people. We used two product sets from different domains to demonstrate its domain-independence: greeting cards and holiday homes. The greeting card set contains Dutch tags, while the holiday homes did not contain any tags, but only descriptions. From the holiday homes, we used the English descriptions to extract (English) tags, to emphasize the potential to use IBRS in a language-independent way.

This section is further structured as follows: Section 5.1 describes the item set details. In Section 5.2, we present the approach taken to validate both our ranking mechanism and the recommendation explanation interface. Section 5.3 finally, discusses the validation results.

5.1 Item set details

The first item set contains greeting cards from the Dutch company *Kaartje2Go* ("Card2Go"). People search through a collection of cards electronically, which are distributed through regular (non-electronic) mail by Kaartje2Go in name of the customer. The customers can choose between sending greeting cards to one or multiple people at once. 75% of the purchases are of the latter type, for which the preferences of the sender are more relevant than those of the (potentially many) recipients. To facilitate the search, users can search for tags that have been entered manually by the Kaartje2Go employees. These tags, which are mostly in Dutch, are inconsistent in their completeness: for example some of the soccer cards are also tagged using the names of popular Dutch soccer teams, but not all of them. Less popular teams are never mentioned as tags. The top-10 of the translated greeting card tags can be found in Table 2.

The second item set contains holiday homes from the holiday home portal *EuroCottage*. This item set did not contain tags, but a description in one, two or three languages (Dutch, English and/or German). We followed the approach discussed in Section 4.3 to extract mentions of geographic places from the English holiday home descriptions. The top-10 of resulting tags can be found in Table 3. The advantage of extracting geographic places is that these also often have Wikipedia pages, which makes them suitable for the requirement that the tags need to have a DBPEDIA_RESOURCE_ID. Many pages of the holiday home descriptions were in German, even though they were entered into the system by the holiday home owners as English descriptions. As a result thereof, many German words or phrases were extracted as

Tag	Frequency
Birthday	7,535
Party	4,200
Love	2,521
Girl	2,268
Boy	2,084
Infant	2,056
Photograph	1,793
Marriage	1,543
Cool	1,381
Animals	1,373

Table 2: Top-10 of (translated) manual greeting card tags with a DBpedia reference, ordered by the number of cards with this tag

geographical references, since the model was trained for English descriptions. However, the impact of these terms was practically zero, as these extracted tags were not matched with an English DBpedia resource.[3] For the validation, the holiday homes were plotted on a map that was zoomed in on Europe, since most holiday homes in the set are located there. A relatively small subset of homes outside Europe could therefore not be displayed on the map, and were removed from the validation set, just as those without a coordinate pair. This coordinate pair was also used for the diversity function: all top-10 holiday homes had to be located at least 250 kilometers away from higher ranked items.

Tag	Frequency
Florence	760
Siena	656
Mediterranean Sea	634
Tuscany	537
Legoland	513
Venice	508
Sotkamo	448
Europe	440
Ardennes	421
Pisa	363

Table 3: Top-10 of extracted tags for holiday homes with a DBpedia reference, ordered by the number of holiday homes with this tag

5.2 Validation approach

Our test users were requested to participate through Facebook, and used their own existing Facebook account for the recommendations. The test users were not aware of *what* they were testing, except for the information that they were testing a RS. Most test users do not have a background in computer science, and none of them were aware of how IBRS works. We asked our test users to validate our algorithm through a total of 30 questions, split up into three batches of 10. Once a question had been answered, users could not return to that question. The first two batches were intended to validate our ranking mechanism, the third batch was intended to determine the user perception of recommendations with explanations, as compared to recommendations without explanations.

For the first ten questions, users were asked to select their favorite greeting card from a greeting card pair using the interface of Figure 3. On one side of the screen, an item from the top-10 greeting cards according to IBRS was shown. On the other side, a card was shown that was not tagged with any of the matched tags. We called these recommendations *Inverted IBRS*. IBRS and Inverted IBRS were shown on the left or right side at random.

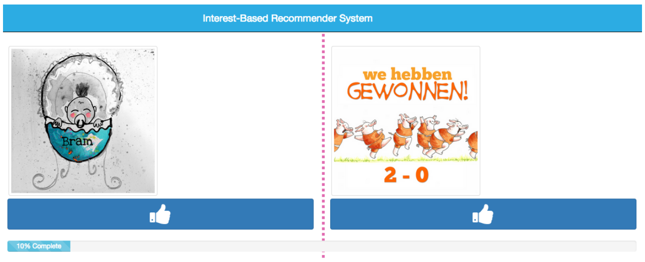
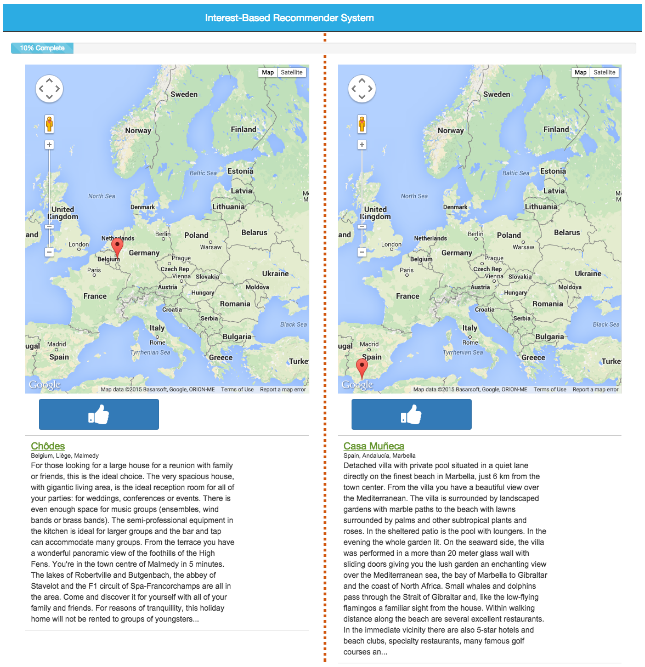


Figure 3: Validation interface for greeting card comparison

For the second batch of ten questions, our test users were presented with the choice between two holiday homes, in a similar way. Again, IBRS and Inverted IBRS were shown on the left or right side at random. For each holiday home, its location was shown on a map, with the name of the holiday home and the first 1000 characters of its description, as shown in Figure 4.



Chôdes
Belgium, Liège, Malmedy
For those looking for a large house for a reunion with family or friends, this is the ideal choice. The very spacious house, with gigantic living area, is the ideal reception room for all of your parties: for weddings, conferences or events. There is even enough space for music groups (ensembles, wind bands or brass bands). The semi-professional equipment in the kitchen is ideal for larger groups and the bar and tap can accommodate many groups. From the terrace you have a wonderful panoramic view of the foothills of the High Fens. You're in the town centre of Malmedy in 5 minutes. The lakes of Robertville and Butgenbach, the abbey of Stavelot and the F1 circuit of Spa-Francorchamps are all in the area. Come and discover it for yourself with all of your family and friends. For reasons of tranquillity, this holiday home will not be rented to groups of youngsters...

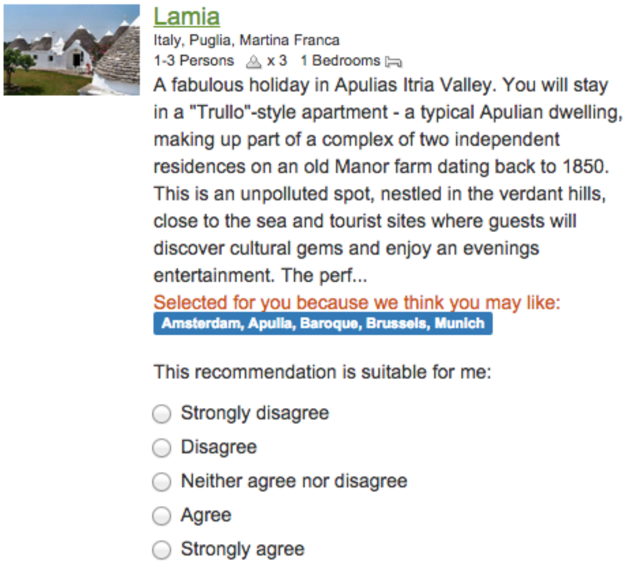
Casa Muñeca
Spain, Andalucía, Marbella
Detached villa with private pool situated in a quiet lane directly on the finest beach in Marbella, just 6 km from the town center. From the villa you have a beautiful view over the Mediterranean. The villa is surrounded by landscaped gardens with marble paths to the beach with lawns surrounded by palms and other subtropical plants and roses. In the sheltered patio is the pool with loungers. In the evening the whole garden lit. On the seaward side, the villa was performed in a more than 20 meter glass wall with sliding doors giving you the lush garden an enchanting view over the Mediterranean sea, the bay of Marbella to Gibraltar and the coast of North Africa. Small whales and dolphins pass through the Strait of Gibraltar and, like the low-flying flamingos a familiar sight from the house. Within walking distance along the beach are several excellent restaurants. In the immediate vicinity there are also 5-star hotels and beach clubs, specialty restaurants, many famous golf courses an...

Figure 4: Validation interface for holiday home comparison

The final batch of ten questions required the test users to rate a recommendation. Each of the holiday homes was one of the top-10 holiday homes according to IBRS. At random, a user was assigned to the group of users who received rec-

[3] Even though the approach can be applied to any language contained in the knowledge base, the tags are still matched with knowledge base resources in the tag language.

ommendations *with* an explanation, as shown in Figure 5, or *without* an explanation.



Lamia

Italy, Puglia, Martina Franca

1-3 Persons x 3 1 Bedrooms

A fabulous holiday in Apulias Itria Valley. You will stay in a "Trullo"-style apartment - a typical Apulian dwelling, making up part of a complex of two independent residences on an old Manor farm dating back to 1850. This is an unpolluted spot, nestled in the verdant hills, close to the sea and tourist sites where guests will discover cultural gems and enjoy an evenings entertainment. The perf...

Selected for you because we think you may like:

Amsterdam, Apulia, Baroque, Brussels, Munich

This recommendation is suitable for me:

○ Strongly disagree

○ Disagree

○ Neither agree nor disagree

○ Agree

○ Strongly agree

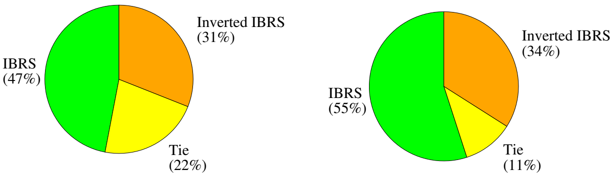
Figure 5: Cut-out of validation interface for holiday home recommendation rating. The lines in orange/blue contain the matched tags.

In test runs of the validation process, we determined that in a set-wise comparison of the two systems, users tended to prefer the set that was spread out over the map, rather than one that contained clusters of recommendations. Since Inverted IBRS is extremely spread out, due to the fact that items had no relation with the users or each other, this caused a bias in the validation results. Therefore, we decided to only compare the results item-wise. Furthermore, we removed tags with a negative connotation, such as “die,” or “death.”

5.3 Validation results

The first two batches of the validation were used to determine the potential of the IBRS ranking mechanism. The results are shown in the pie charts of Figure 6. Figure 6a shows which system was the test user’s preferred system, based on a majority vote between the two systems. Most users participated in the validation of both the recommendation of greeting cards and holiday homes. Each batch was counted separately. 47% of the users preferred IBRS, 22% voted equally often for both of the systems, and 31% of the users preferred Inverted IBRS. In the pie chart of Figure 6b, the results are shown when the results of holiday homes with the greeting cards are combined per user. Since this increases the number of votes per user, ties are less common. In this scenario, 55% of the users preferred the IBRS results, while 34% preferred Inverted IBRS.

The final batch of the validation was used to determine the usefulness of the proposed recommendation explanation interface for holiday homes. The results of this batch are shown in the histograms of Figure 7. Contrary to our expectations, users preferred to receive recommendations without explanations. Using the 5-point Likert scale, the users who were presented with an interface with explanations rated the recommendations with an average score of 3.3772, while users without recommendation explanation rated the recommendations with a 3.4709 on average. From this validation, we can conclude that people that receive recommendations based on tags that do not describe them well, are more likely to reject a recommendation with a “strongly disagree,” when they see the rationale behind the recommendation.

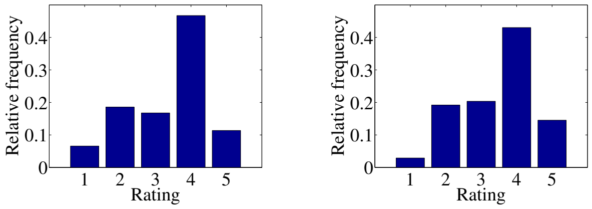


(a) Split out between greeting cards and holiday homes (batches counted separately)

(b) Overall (batches combined)

Figure 6: Most frequent choices per user for the first two batches of questions

Despite satisfying results with respect to the system’s potential to rank recommendations for users, we should not forget that many aspects play a role in the decision-making that cannot (yet) be detected from Facebook profiles. When choosing either a greeting card, a holiday home, or anything else, one will always look at domain-specific item characteristics. For a greeting card, the user looks at colors, style, and the occasion the card is sent for. Similarly, for a holiday home, he looks at price, number of beds, the picture of the home, and the distance to the beach. For this reason, this approach shall only be used as a feature of a larger system.



(a) With recommendation explanation; average rating: 3.3772.

(b) Without recommendation explanation; average rating: 3.4709.

Figure 7: Recommendation ratings split out by recommendation presentation interface

6. CONCLUSION

In this paper, we presented the approach behind IBRS. We discussed the concept of mapping items marked as preferred or liked in social media onto a generic knowledge-base, and query expansion using DBpedia. We presented the technology, including the abstraction layer, tag generation approach, and ranking mechanism. We also presented the validation results of a test user group. As said, we recommend to use the proposed and validated approach from this paper as a feature of a larger recommender system. In a more complete system, one also needs to take domain-specific features, as well as item popularity and other collaborative filtering features, into account. However, these features would contradict with our objective to create a generic RS that overcomes the cold-start problem, and therefore were not taken into account in this work.

Currently, IBRS uses all paths in the knowledge base graph as an indication for a useful recommendation. However, some paths in the graph actually form a reason *not* to recommend that item. For example, in the holiday home domain, a user is less likely to book a home in his own town, even though there may be many paths between him and that holiday home based on his local likes. Furthermore, some nodes are more useful than other for recommendation. DBpedia nodes like "European Central Time" have a lot of incoming paths, while it is unlikely that this actually forms an interest for this user. The next step for IBRS is to further improve the ranking mechanism by incorporating these characteristics and explore the possibility to automatically detect (negative) weights of paths.

7. ACKNOWLEDGEMENTS

This publication was supported by the Dutch national program COMMIT/. We also thank Mena Habib for his support in the tag generation process.

8. REFERENCES

[1] Facebook, "Facebook | photos." `https://www.facebook.com/facebook`, 2013.

[2] S. Bakers, "Statistics of the top facebook pages." `http://www.socialbakers.com/statistics/facebook/pages/total/`, 2013.

[3] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," *Knowledge-Based Systems*, vol. 46, pp. 109–132, 2013.

[4] D. Fijalkowski and R. Zatoka, "An architecture of a web recommender system using social network user profiles for e-commerce," in *Computer Science and Information Systems (FedCSIS), 2011 Federated Conference on*, pp. 287–290, IEEE, 2011.

[5] I. Guy, N. Zwerdling, I. Ronen, D. Carmel, and E. Uziel, "Social media recommendation based on people and tags," in *Proc. of the 33rd intern. ACM SIGIR conference on Research and development in information retrieval*, pp. 194–201, ACM, 2010.

[6] J. He and W. W. Chu, *A social network-based recommender system (SNRS).* Springer, 2010.

[7] A. Passant, "dbrec - music recommendations using DBpedia," in *The Semantic Web–ISWC 2010*, pp. 209–224, Springer, 2010.

[8] A. Passant and Y. Raimond, "Combining social music and semantic web for music-related recommender systems," in *The 7th International Semantic Web Conference*, p. 19, Citeseer, 2008.

[9] R. Mirizzi, T. Di Noia, A. Ragone, V. C. Ostuni, and E. Di Sciascio, "Movie recommendation with DBpedia," in *IIR*, pp. 101–112, Citeseer, 2012.

[10] J. Golbeck and J. Hendler, "Filmtrust: Movie recommendations using trust in web-based social networks," in *Proceedings of the IEEE Consumer communications and networking conference*, vol. 96, pp. 282–286, University of Maryland, 2006.

[11] B. N. Miller, I. Albert, S. K. Lam, J. A. Konstan, and J. Riedl, "MovieLens unplugged: experiences with an occasionally connected recommender system," in *Proceedings of the 8th international conference on Intelligent user interfaces*, pp. 263–266, ACM, 2003.

[12] V. C. Ostuni, T. Di Noia, R. Mirizzi, D. Romito, and E. Di Sciascio, "Cinemappy: a context-aware mobile app for movie recommendations boosted by DBpedia," *SeRSy*, vol. 919, pp. 37–48, 2012.

[13] P. Symeonidis, A. Nanopoulos, and Y. Manolopoulos, "Moviexplain: a recommender system with explanations," in *Proceedings of the third ACM conference on Recommender systems*, pp. 317–320, ACM, 2009.

[14] B. Heitmann and C. Hayes, "Using linked data to build open, collaborative recommender systems.," in *AAAI spring symposium: linked data meets artificial intelligence*, pp. 76–81, 2010.

[15] T.-P. Liang, Y.-F. Yang, D.-N. Chen, and Y.-C. Ku, "A semantic-expansion approach to personalized knowledge recommendation," *Decision Support Systems*, vol. 45, no. 3, pp. 401–412, 2008.

[16] S. Bostandjiev, J. O'Donovan, and T. Höllerer, "TasteWeights: a visual interactive hybrid recommender system," in *Proc. of the 6th ACM conf. on Recommender systems*, pp. 35–42, ACM, 2012.

[17] R. Burke, "Hybrid web recommender systems," in *The adaptive web*, pp. 377–408, Springer, 2007.

[18] P. Lops, "Semantics-aware content-based recommender systems," 10 2014. Keynote at Workshop on New Trends in Content-based Recommender Systems.

[19] P. Basile, C. Musto, M. de Gemmis, P. Lops, F. Narducci, and G. Semeraro, "Content-based recommender systems + DBpedia knowledge = semantics-aware recommender systems," in *Semantic Web Evaluation Challenge*, pp. 163–169, Springer, 2014.

[20] C. Shi, C. Zhou, X. Kong, P. S. Yu, G. Liu, and B. Wang, "HeteRecom: A semantic-based recommendation system in heterogeneous networks," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1552–1555, ACM, 2012.

[21] C. Shi, X. Kong, Y. Huang, S. Y. Philip, and B. Wu, "HeteSim: A general framework for relevance measure in heterogeneous networks," *IEEE Transactions on Knowledge & Data Engineering*, no. 10, pp. 2479–2492, 2014.

[22] V. Zanardi and L. Capra, "Social ranking: uncovering relevant content using tag-based recommender systems," in *Proceedings of the 2008 ACM conference on Recommender systems*, pp. 51–58, ACM, 2008.

[23] M. B. Habib and M. van Keulen, "Improving toponym disambiguation by iteratively enhancing certainty of extraction," in *Proceedings of the 4th International Conference on Knowledge Discovery and Information Retrieval, KDIR 2012, Barcelona, Spain*, (Spain), pp. 399–410, SciTePress, October 2012.

[24] DBpedia, "SPARQL explorer for http://dbpedia.org/sparql." `http://dbpedia.org/snorql/`, 2015.

[25] V. de Graaff, M. van Keulen, and R. A. de By, "Towards geosocial recommender systems," in *4th Intern. Workshop on Web Intelligence & Communities (WI&C 2012), Lyon, France*, ACM, 2012.