Int J Digit Libr
DOI 10.1007/s00799-017-0224-8

Retrieval by recommendation: using LOD technologies to improve digital library search

Lisa Wenige[1] · **Johannes Ruhland**[1]

Received: 17 October 2016 / Revised: 3 July 2017 / Accepted: 11 July 2017

Abstract This paper investigates how Linked Open Data (LOD) can be used for recommendations and information retrieval within digital libraries. While numerous studies on both research paper recommender systems and Linked Data-enabled recommender systems have been conducted, no previous attempt has been undertaken to explore opportunities of LOD in the context of search and discovery interfaces. We identify central advantages of Linked Open Data with regard to scientific search and propose two novel recommendation strategies, namely *flexible similarity detection* and *constraint-based recommendations*. These strategies take advantage of key characteristics of data that adheres to LOD principles. The viability of Linked Data recommendations was extensively evaluated within the scope of a web-based user experiment in the domain of economics. Findings indicate that the proposed methods are well suited to enhance established search functionalities and are thus offering novel ways of resource access. In addition to that, RDF triples from LOD repositories can complement local bibliographic records that are sparse or of poor quality.

Keywords Recommender systems · Linked (Open) data · Information retrieval · Discovery interfaces · Knowledge Organization Systems

1 Introduction

As of 2014, open bibliographic data comprised 9.47% of the entire Linked Open Data (LOD) cloud with millions of triple statements spread over 87 datasets.[1] But up to this date, only a few applications make active use of these valuable data sources. While there is extensive research on Linked Data-enabled Recommender Systems (LDRS) for decision support in online shops (see Sect. 2), opportunities of LOD have not yet been thoroughly investigated for scientific publication retrieval, even though recommendation features are part of most digital library and discovery interfaces [45]. Various recommendation strategies, that incorporate bibliographic, user and/or citation data, have shown to produce useful suggestions [12]. In addition to that, recent evidence has proven that recommendations can enhance general search through offering a low-effort means of resource discovery, especially when users are not familiar with a domain or when they have difficulties formulating a query [41]. Linked Open Data, on the other hand, can help digital libraries to overcome issues of data sparsity and data quality. As we will show in this paper, recommendations from SPARQL endpoints can be calculated on-the-fly and therefore can be seamlessy integrated into digital library interfaces. Thus, LOD might give way to novel ways of resource access by combining general search queries and recommendation methods. For this purpose, SKOS annotations of LOD resources can be evaluated.

Figure 1 depicts bibliographic RDF descriptions for an example LOD resource (`econstor:21555`). The RDF graph represents a publication from the Open Access repository EconStor and a subset of its metadata. Among general metadata information, such as the title (`rdfs:label`) and the language (`dc:language`) of the publication, the RDF graph contains annotations that stem from the SKOS version of the Standard Thesaurus Economics [20].[2] These annotations contain valuable information that can be used

✉ Lisa Wenige
lisa.wenige@uni-jena.de

[1] Chair of Business Information Systems, Friedrich-Schiller-University Jena, Jena, Germany

[1] http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state/.

[2] https://www.econstor.eu/.

Published online: 19 July 2017

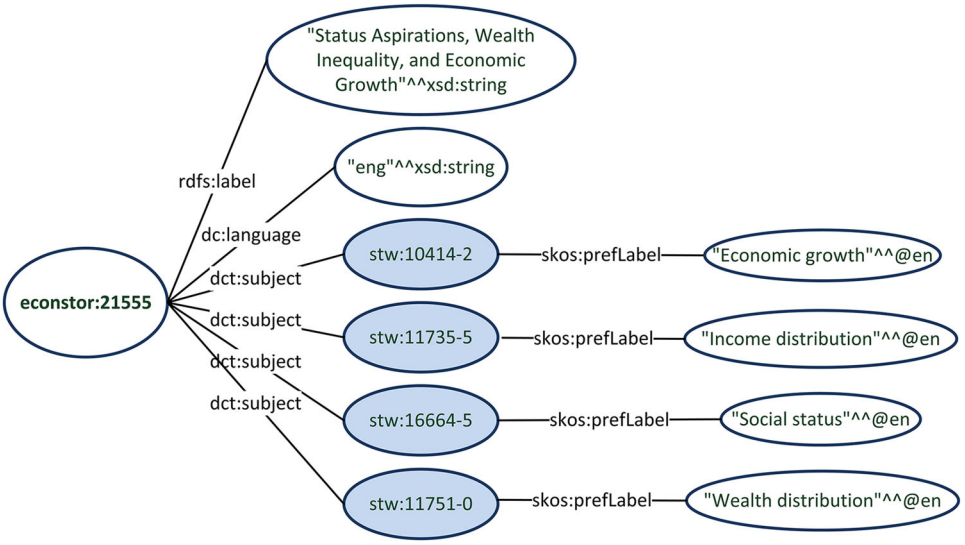

Fig. 1 Bibliographic RDF descriptions for an example resource of the LOD repository EconStor

for similarity calculation during recommendation retrieval. Based on this example we will show, how SKOS annotations and Linked Open Data can be used for automatic suggestions.

In summary, the central contributions of the paper are:

- A comprehensive survey of the possibilities of LOD for recommendations and information retrieval within the digital library context.
- Development of an approach to generate on-the-fly recommendations for scientific publications from remote SPARQL endpoints.
- Refinement of existing topic expansion methods through the application of graph-based similarity metrics.
- Development of a method to combine similarity calculation with expressive SPARQL graph patterns, such that digital library users can effectively filter recommendations according to their needs.

The outline of the paper is as follows: Sect. 2 summarizes related work on recommender systems, LDRS and Linked Open Data. In the subsequent section (Sect. 3), we describe in detail advantages and modes of application for Linked Data-enabled recommendation and search tasks. In this context, two novel retrieval strategies are introduced. Section 4 outlines the setup of our web-based experiment, which was followed through to evaluate the proposed recommendation strategies. The following sections analyze experimental results (Sect. 5) and summarize the most important findings (Sect. 6).

2 Related work

Since the publication of the first papers in the 1990s, recommender systems (RS) have been an extensively studied field. Due to their status as a key technology in major e-commerce applications, RS research is often driven by application-specific demands [29]. In the digital library context, the extent of digital collections and diversified information needs made effective user support necessary [64]. Thus, a variety of content-based (CB) [48], collaborative filtering (CF) [30,50] and hybrid systems [47] were developed. Additional methods, such as citation analysis [49] or taxonomic topic expansion [47] were introduced to enhance CF algorithms. In the area of information retrieval and resource access, additional systems exist as an add-on for Online Public Access Catalogues (OPACs). In these systems preferences are inferred from user interest histories, that either come from loan records [64], session data [27] or explicit user feedback [65].

But feedback information tends to be sparse, which often affects recommendation quality [29]. Hence, other authors proposed to evaluate co-occurrences in session data in order to identify related publications [27]. But with this approach, the central disadvantage is that information needs might differ widely, especially when users search for literature in specialized scientific domains. Thus, content-based recommendation strategies are a good alternative in the retrieval context. These strategies have already been explored by other researchers. But while the problem of missing overlapping user interests does not apply here, metadata descriptions might also be insufficient [12,47,48].

For domains other than digital libraries, such as tourism or multimedia, some authors proposed to apply information from Linked Open Data repositories [58,59] to address problems of data sparsity and quality. Many followed this proposal and utilized LOD for content-based recommendations [31,36,51,53–58].

As libraries and other information providers are among the institutions that most actively participate in the LOD movement, they would be predestined to use open data

sources to enhance their services. Since the publication of the W3C Recommendation for Simple Knowledge Organization Systems (SKOS), numerous thesauri (e.g., the Library of Congress Subject Headings [LCSH], the STW Thesaurus for Economics [STW], the Thesaurus for the Social Sciences [TheSoz] or the AGROVOC Multilingual Agricultural Thesaurus [AGROVOC]) were published [20,60,63,67]. Additional efforts concerned the creation of mappings between controlled vocabularies [68] and the publication of authority files and bibliographic records from local or union catalogues [62]. By now, most of these activities are coined under terms of Linked Library Data (LLD), which are further accompanied by networking bodies, such as the W3C Library Linked Data Incubator Group.[3]

But despite this active involvement in LLD activities, yet only a few applications exist, that make use of LOD for information retrieval [34,61]. As opposed to RS research in other domains, which has produced a variety of LDRS, there has been no comparable effort for scientific publications. To the best of our knowledge, we are the first to investigate the opportunities that LLD data has to offer for recommendations in information retrieval.

3 Linked library data for recommendation

3.1 Openness

Over the past years, many libraries and information centers have opened up their catalogue data to the general public by adhering to best practices of the Linked Open Data movement [62]. These principles entail aspects of open publishing. Documents are required to be published as URI resources, that can be either accessed via HTTP or SPARQL endpoints [26]. As such, (digital) libraries are not dependent on their local data sources, but can enrich their services with additional metadata from the same or a related domain. More than 85% of these publicly accessible bibliographic datasets are registered on the open data CKAN repository and can be easily discovered.[4] As (digital) libraries often face tight budgets, Linked Open Data might offer a chance to enrich services without imposing too much additional costs.

But even though open bibliographic data offers promising opportunities, only a few providers explicitly state terms under which the data can be used and modified.[5] In the future, data publishers should provide comprehensive licensing information (e.g., by using a license from the Open Data Commons (ODC) license framework) to assure data consumers that they are not infringing any rights [26].

[3] http://www.w3.org/2005/Incubator/lld.

[4] https://datahub.io/de/dataset?q=lld.

[5] http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state.

3.2 Independence of local data

Besides their public availability, bibliographic datasets have the advantage of being independent from data sources of a specific (digital) library. Local user and publication metadata do not necessarily need to be accessed to provide interesting recommendations as long as there are suitable records from the LOD cloud that can be exploited. Thus, privacy issues do not occur as no user data are processed. While adherence to privacy legislation is by all means necessary, it still sets some boundaries that could be overcome by third-party datasets [27].

Additionally, the performance of recommender systems depends to a large extent on the available data sources. Hence, Linked Open Data could be of value when the amount or quality of local records is insufficient. For instance, it has often been stated, that collaborative filtering systems perform poorly when rating data is sparse. This is quite often the case, because users avoid the extra work to provide explicit feedback [3,28–30]. A possible workaround would be to apply implicit feedback, such as page view time or download statistics. These information can be obtained by processing session logs. But in the digital library context, information needs among researchers might be especially diverse, which further reduces access frequency.

In contrast, content-based algorithms only need bibliographic information to provide users with personalized recommendations. But even when following this approach, metadata might be missing due to insufficient indexing. Here, datasets from the Linked Library Data cloud (see Sect. 2) could succeed in providing additional data for matching entities (e.g., publications, authorities, subjects). Usually library data tends to be of high quality as cataloguing is carried out by trained professionals [62].

3.3 On-the-fly recommendations

RDF triples on the Web of Data reflect the status of their latest update and can be used when actually needed, as Bizer put it:

> The Web of Data is open, meaning that applications do not have to be implemented against a fixed set of data sources, but can discover new data sources at runtime by following RDF links [26].

Thus, resources can be retrieved on-the-fly. As long as a publication (e.g., a book or a scientific paper) is uniquely identified through a URI and is annotated with concepts from a SKOS vocabulary, LOD can be utilized to generate *ad hoc* recommendations through an API-like interface. SPARQL endpoints could serve as primary access sources, from which suitable resources are detected.

The principles of on-the-fly recommendation were introduced in [32], but will be briefly summarized to ease understanding. Our recommendation engine generates suggestions based on SKOS annotations of an input publication r. Thereby the engine looks at the annotation graph (AG) of an RDF dataset, that solely contains triple statements of publications and their SKOS annotations. Publications and annotations are assumed to be connected with a standard annotation property, e.g., as proposed by the Metadata Terms of the Dublin Core Metadata Initiative (DCMI) (e.g., `http://purl.org/dc/terms/subject`, abbr. `dct:subject`). The recommendation engine issues a SPARQL query solely containing P_r as a basic graph pattern in the where-condition against a SPARQL endpoint to determine all publications that have at least a single annotation in common with publication r (Eq. 1). The SPARQL server evaluates the pattern (P_r) against the annotation graph (AG) in its repository and returns all publications and their SKOS annotations that fit the given condition (Eq. 2).

$$P_r = (r, dct:subject, ?c)\ AND\ (?x, dct:subject, ?c) \tag{1}$$

$$\Omega_r = [\![P_r]\!]_{AG} \tag{2}$$

The solution from this query is a multiset (Ω_r), that can be used for content-based recommendation retrieval. For each publication $q \in \Omega_r$, similarities are calculated based on the information content (IC) of the set of its shared features (C) with input resource r (Eq. 3). IC is defined as the sum of the individual information content values of each concept that is contained in the set of shared features ($c \in C$) (see Eq. 4) and where $freq(c)$) is the frequency of c among all relevant resources ($q \in \Omega_r$) and n is the maximum frequency among these resources. The final similarity score of two resources r and q is determined by summating the IC values of each concept of the shared feature set (Eq. 5).

$$C = \{Annot(r) \cap Annot(q)\} \tag{3}$$

$$IC(C) = -\sum_{c \in C} log\left(\frac{freq(c)}{n}\right) \tag{4}$$

$$score(r, q) = IC(C) \tag{5}$$

Figure 2 illustrates the previous definitions. For instance, it is known that a digital library user showed interest (e.g., through full text access or download) in the EconStor publication `econstor:21555`, for which metadata descriptions are available in the LOD cloud. The recommendation engine takes this resource and looks for other resources with matching SKOS annotations through accessing a SPARQL endpoint. The more an LOD resource has in common with the input paper, as indicated by many matching resources with high information content (IC) values, the higher they are ranked during recommendation retrieval. In the given example, resource `econstor:21555` is found to be related to resource `econstor:34662`, since the two publications share two annotations with comparably high information content values (`stw:11735-5` and `stw:11751-0`) (see Fig. 2).

But since many bibliographic RDF databases contain snapshots of entire library catalogues and certain concepts are frequently used to annotate publications, the recommendation engine has to process numerous triple statements. This can lead to increased computational times. Thus, we intend to reduce processing effort through introduction of a lower

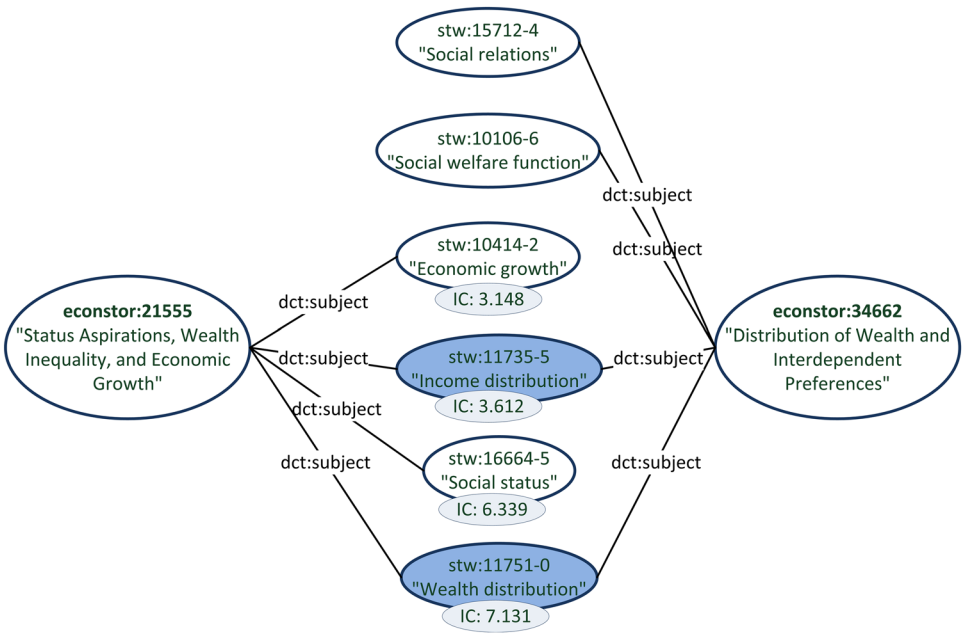

Fig. 2 SKOS concept matching of two LOD resources for similarity calculation (on-the-fly recommendation retrieval)

bound for concept annotations. Based on the preferred number of suggestions (k) (e.g., as specified by the digital library user) and the potential IC values, the system precalculates the minimum number of annotations that a publication has to share with an input publication. This condition is met as soon as the maximum potential information content for a given number of matching annotations is smaller than the minimum potential score for a higher number of shared features. Thus, many resources can be omitted without affecting the final ranking. This approach has already been tested on various RDF datasets and proved to considerably reduce processing times [32].

One of its central advantages, as opposed to previous Linked Data-enabled recommender systems (see Sect. 2), is its independence of any processing steps with regard to natural language computation or physical data integration. It enables even small digital libraries or such with limited funding to make the most of all publicly available bibliographic records in their domain.

3.4 Flexible similarity detection

As opposed to the simple similarity approach based on exact matching of annotations, we assume that enhancing annotations with similar concepts helps to better explore document collections. It mirrors the common retrieval technique, where a complex information need is translated into a query by representing topical concepts through search terms that are connected by OR expressions. In turn, these topical concepts are linked by AND operators [33]. For instance, an economist, who does research on the connection of wealth equality and economic growth would pose the following query:

("Income distribution" OR "Wealth distribution")
AND
("Economic growth" OR "Economic policy")

This technique is known as (automatic) query expansion and has a long tradition in information retrieval systems. Here, search terms are enhanced with additional terms that are similar to the terms of the original query in order to increase recall. Similar terms can be identified with various techniques. A natural starting point would be to explore linguistic variations that are based on the word stem of the query term. Another technique determines related terms on the basis of term co-occurrence and proximity analysis from a text corpus. Additionally, as soon as a document is annotated with concepts from a controlled vocabulary, query expansion can draw upon the semantic relationships of the thesaurus in order to detect related concepts that can be of use in the context of the query. This technique has been applied in many existing retrieval systems, that use local data sources [33]. Openly accessible SKOS vocabularies have been recently adopted for query reformulation as well [34,35].

In concordance with that, we consider SKOS-based topic expansion to be a viable method for recommendation tasks. First evidence for this hypothesis was provided by the authors of [31] and [36]. They enhanced subject annotations with related concepts from `skos:broader` connections in their LDRS, which showed high accuracy values.

In [37], researchers developed this approach one step further by quantifying relatedness of annotations through a taxonomic similarity measure. By doing so, they made the justified claim that similarity depends on the position of both concepts in a taxonomy and that for content-based recommendations to be effective, relatedness should be based on their specificity, i.e., hierarchy level. But in their recommendation experiment, the related terms expansion method based on concept similarity did not outperform the non-expansion method. This might be due to the fact that related terms were considered as being equally important as input profile annotations.

Based upon these findings, we incorporate taxonomically related concepts while accounting for their decreased similarity values. Hence, we propose to apply common graph-based similarity metrics for concept expansion (e.g., by [19]), instead of only exploring direct annotations as it has often been done in existing retrieval and recommender systems [31,33,36].

We define the set of proximate concepts $Prox(a, r)$ for an annotated concept a of a publication r in the user profile as in Eq. 6. These proximate descriptors need to exceed a threshold level t of similarity with concept annotation a. Even though, concept expansion methods impose some overhead in terms of computational complexity, digital library providers would be able to avoid long processing times due to increased result set sizes by setting t sufficiently high (e.g., $t > 0.75$).

$$Prox(a, r) = \{\ p \mid a, p \in AG,\ a \in Annot(r),\ s(a, p) \geq t\ \} \tag{6}$$

In order to execute the recommendation procedure in concept expansion mode, the system needs to get hold of the union of all similar descriptor sets of each annotation a of the input publication r (see Eq. 7).

$$Prox(r) = \bigcup_{a \in Annot(r)} Prox(a, r) \tag{7}$$

These similar descriptors can then be applied for the retrieval of potentially relevant resources from an LOD repository. The engine extracts all publications that are annotated with concepts ($?c$) (see Eq. 8), which are contained in the set of all proximate descriptors of r ($\mu(?c) \in Prox(r)$)

or do have matching annotations with the input resource (Ω_r) or share both exact and similar matches with it (see Eq. 9).

$$P_f = (?x, dct : subject, ?c) \tag{8}$$

$$\Omega_f = \{\ \mu \mid \mu \in [\![P_f]\!]_{AG},\ \mu(?c) \in Prox(r)\ \} \cup \Omega_r \tag{9}$$

Upon having extracted all potentially relevant publications, similarity values have to be determined for each of these resources with regard to the input publication. The similarity score of a potentially relevant publication q with input publication r is calculated by comparing the features of these two resources. For each annotation a of the input resource r the engine evaluates, whether it is contained in the annotation set of resource q. When there is a matching, the IC value of a is used for similarity calculation (see Eq. 10). In case, the concept is not contained in the annotation set of resource q, but there are similar concepts of a in $Prox(a, r)$ that are also contained in the annotation set of q, then their concept-to-concept similarity values are stored in the set $S(a, r, q)$ (see Eq. 11), from which the maximum value is selected (see Eq. 12). The value is then used to weight the IC value of a and thus to slightly decrease the score. This is done in order to account for the fact that for this particular annotation, there only exists a similar matching. In case of the two publications r and q sharing neither an exact nor a proximate match, the score value is set to zero (see Eq. 10).

$$d(a, r, q) = \begin{cases} IC(a), & a \in Annot(q) \\ sim(a, r, q) \times IC(a), & a \notin Annot(q), S(a, r, q) \neq \emptyset \\ 0, & otherwise \end{cases} \tag{10}$$

$$S(a, r, q) = \{\ s(a, p) \mid p \in Prox(a, r) \cap Annot(q)\ \} \tag{11}$$

$$sim(a, r, q) = max(S(a, r, q)) \tag{12}$$

Upon having obtained the score values of all annotations for a given user profile publication r and a potentially relevant publication q, the final similarity value of these two LOD resources is determined by the sum of the scores for each annotation of r (see Eq. 13).

$$score_f(r, q) = \sum_{a \in Annot(r)} d(a, r, q) \tag{13}$$

For illustration purposes, consider the example given in Fig. 3. Here, similarity calculation is not only based on matching annotations, but on similar annotations as well. In this example, concepts are considered to be related as soon as their similarity value, as measured by a graph-based similarity metric, exceeds the given threshold ($t > 0.75$). Even though the input paper (`econstor:21555`) does not share the SKOS concept "Wealth distribution" (`stw:11751-0`) with the potentially relevant LOD resource in this example (`econstor:34662`), the related concept "Savings" is taken into account for recommendation retrieval. This is because the precalculated concept-to-concept similarity value for "Wealth distribution" and "Savings" is sufficiently high ($sim = 0.816$). The concept "Overlapping generations" (`stw:10179-0`) is considered to be related as well, but is not taken into account, because its similarity with the concept "Wealth distribution" is smaller than the similarity value of the "Savings" concept. Additionally, the fact

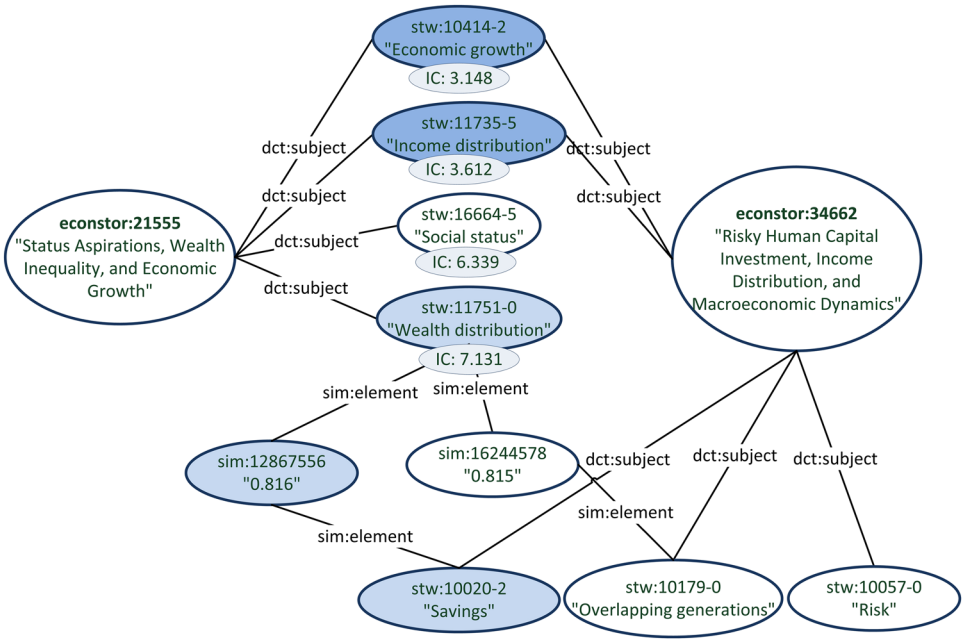

Fig. 3 Expansion with similar SKOS concepts while matching annotations (flexible similarity calculation)

that the SKOS concepts "Wealth distribution" and "Savings" do not match exactly is taken into account through multiplication of the IC value of "Wealth distribution" with the concept-to-concept similarity of both "Wealth distribution" and "Savings" (0.816×7.131). Thus, related concepts are contributing less to the final similarity score of the two resources than matching concepts.

The presented flexible similarity calculation method allows for *ad hoc* recommendations based on a given profile, that contains one or more previously liked publications. Thus, digital library users do not need to set up a separate account, but can receive suggestions on-the-fly from session data. As only concept-to-concept similarities need to be precalculated, users can browse among different granularity levels of concept-to-concept similarity (i.e., t set to different values) until the results best fit their needs.

3.5 Constraint-based recommendations

User behavior has been extensively studied in the context of library search interfaces [40–44]. But while subject indexing potentially offers powerful techniques for retrieval access, users often take not full advantage of them. As is done when using web search engines, queries are formulated as typical keyword searches with limited to no usage of boolean operators [42,44]. During the past years, discovery interfaces were set up to help users navigate through large collections. Among other tools, such as keyword search boxes or faceted search, recommendations are the most widely used features in these next generation catalogues [45]. According to the authors of [41], automatic suggestions are vital to the retrieval process. In an extensive user study they found that recommendations were seen as the most helpful feature for complementing general search functionality. In addition to that, the researchers identified a general tendency of users to turn to the recommendation section from general search. But what is even more striking, is their finding that participants seemed to stick to the recommendation paradigm on chains of suggestions once they found a relevant publication, rather than refining their previous query. This preference seemed to be more prevalent for undergraduate students, which led the researchers to the conclusion that recommendation features in (digital) library catalogues would need to better adapt to advanced user requirements [41].

These findings suggest that recommendations could play an even more prominent role for information retrieval than they currently do. They represent a low-threshold search strategy without requiring end users to formulate complicated queries and hence offer publication access through a retrieval-by-example strategy. As research indicates that users perceive simple keyword searching as well as recommendations to be useful for finding relevant publications, a combination of both paradigms might be of high value.

Linked Library Data could enable this fusion through its expressive data model. Thus, filter conditions from users could not only be used on immediate publication attributes, as it is currently done in faceted search, but could rather be expressed as powerful SPARQL graph patterns P on an openly accessible bibliographic dataset, e.g., incorporating related descriptors for a given subject condition. Eq. 14 shows, how these considerations can be implemented. Before similarity calculation, the user filter is applied through identification of all LOD resources of a dataset D that satisfy a given user condition. These resources are then joined with all relevant resources Ω_r (as defined by Eq. 2).

$$\Omega = \{\mu(?x) | \mu \in [\![P]\!]_D\} \bowtie \Omega_r \tag{14}$$

Consider the example that is given in Fig. 4: A user is known to have accessed a publication on wealth inequality and income distribution, but states that he/she is only interested in publications that reflect on these topics for the region of "Northern Europe" (`stw:16990-6`). Our system translates this request into the below listed SPARQL query and expands the given user filter with subcategories and related categories of the concept "Northern Europe" from the STW thesaurus. Hence, for this example, the graph pattern P is represented by the expression given in the WHERE section of the depicted SPARQL query.

```
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX stw: <http://zbw.eu/stw/descriptor/>

select * where {
{ ?item dc:subject ?subConcept .
?subConcept skos:broader* stw:16990-6. }
UNION { ?item dc:subject ?subConcept .
?subConcept skos:related stw:16990-6.  }
}
```

Through this approach, the engine is able to retrieve similar publications that are not annotated with the specified concept, but still fit the user condition. Thus, any SKOS annotation that has the filter condition as its parent concept would be retrieved. In the given example, the system would identify resources about "Sweden" (`stw:16989-5`) as well, since "Sweden" is a subcategory of "Northern Europe". Once the filter condition has been applied, the similarity calculation process is started on the reduced set of potentially relevant resources. Note that, because the size of the item set has been diminished by the user filter, the IC values of SKOS annotations differ from the IC values of the previous examples (compare Figs. 2 and 3 with Fig. 4).

As is done in the given example, *constraint-based recommendations* can be combined with the *flexible similarity calculation* method in order to increase recall. This approach can then be further complemented by a powerful user interface to ease the selection process (see Sect. 4.2).

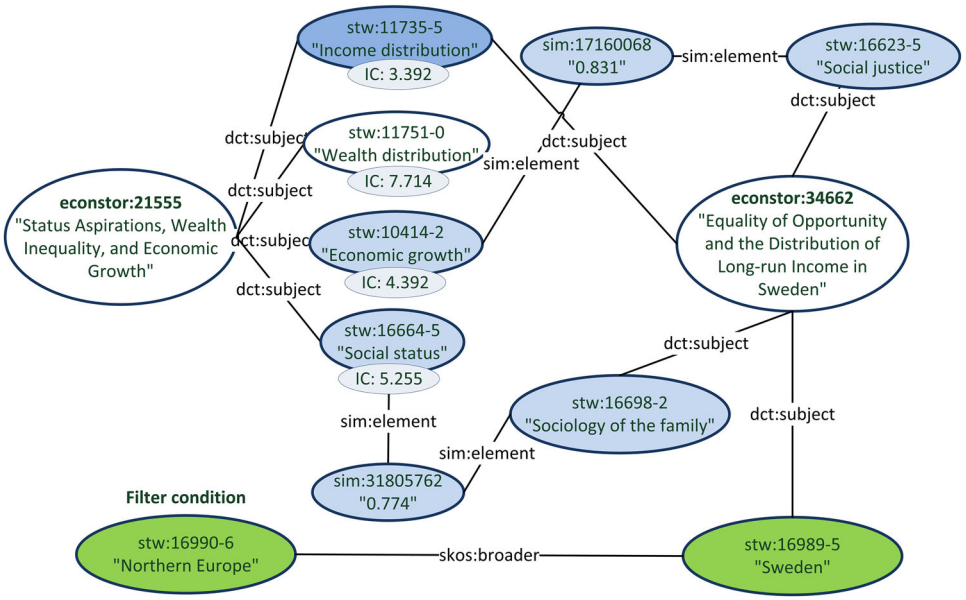

Fig. 4 Application of advanced user filters while matching annotations (constraint-based recommendation retrieval)

4 Evaluation

The effectiveness of the approaches for LOD-enhanced recommendations in digital libraries were evaluated in the context of a many-faceted user study. The following research questions guided our experiment.

4.1 Research questions

R_1. *What constitutes recommendation quality in the digital library context?* There is still a research gap regarding the underlying factors of recommendation quality. Most evaluations of existing recommender systems have a strong focus on *accuracy*. Authors within the research paper RS community [12] and the wider RS community [3,6] argue that this focus does not entail all aspects of a recommender system's performance. It has been pointed out that additional dimensions can have an impact on recommendation quality. In the area of user interaction, issues such as interface design and preference elicitation [9,10] or data security and privacy also influence user satisfaction [1,2,11]. While these aspects should be taken into account when evaluating recommender systems on a holistic level, in our experiment we focused on quality aspects that measure immediate algorithmic performance.

Alongside *accuracy*, it is assumed that *novelty* [3–5] or *diversity* [6–8,13] also play a vital role as quality dimensions. With our evaluation we seeked to investigate, how these dimensions affect overall usefulness in order to get a comprehensive understanding of our recommendation approaches for all subsequent research questions ($R_2 - R_4$).

R_2. *How do text recommendations perform compared to LOD recommendations?* The proposal to use Linked Library Data to enhance search features in digital libraries builds on the assumption that (SKOS) concept annotations are not producing fundamentally worse results than common text-based recommendations.

R_3. *Does flexible similarity detection help to achieve better recommendation results?* The flexible similarity detection method introduced in Sect. 3.4 was analyzed through this research question. It was intended to find out if generating recommendations from different granularity levels of similarity result in different levels of user satisfaction accordingly.

R_4. *Does the possibility to adapt recommendations improve recommendation quality?* This research question seeks to investigate if LOD-enabled user filtering (as it was described in Sect. 3.5) improves performance and how different quality dimensions are affected through user adaptations.

4.2 Online web experiment

Many research paper RS have been validated through offline performance evaluations on datasets. These datasets contain known user feedback for publications. While being the most frequently used method, offline evaluations have the weakness of only measuring performance indicators (e.g., precision or recall) instead of actual satisfaction from real users [3]. It has been shown that performance improvements that were indicated by offline studies did not show the same effect in a related user study [15]. Another drawback of user studies is that they are limited by the sparsity of the datasets, where usually only a few items have been rated by users [3]. Above that, to the best of our knowledge there does not exist any bibliographic LOD dataset, that contains user ratings. Hence, LOD-enhanced recommendation

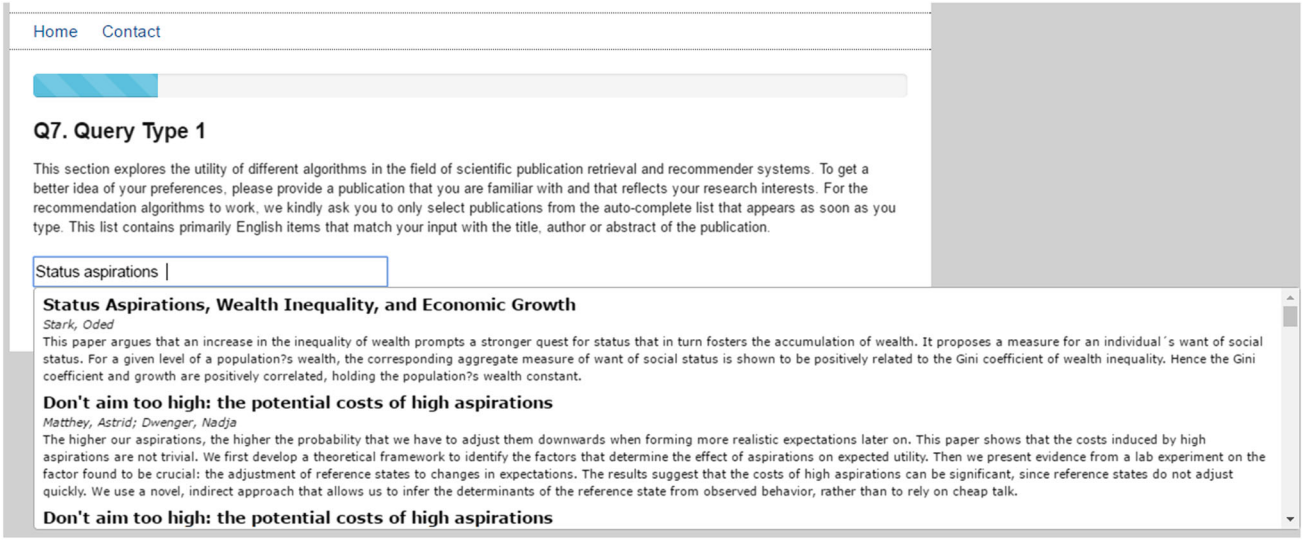

Fig. 5 Web interface for generating a user profile (**Part II** and **Part III**)

approaches for research papers only allow performance evaluations through a user study.

User evaluations can be either carried out as a laboratory experiment or as an online study on a running real world system. Laboratory experiments have the drawback of creating a somewhat artificial setting, where users are aware of being observed, which can bias evaluation results. Studies on existing systems, on the other hand, allow for realistic observations, but might not be available [12]. While we did not have the possibility to perform evaluations on a real world system, we tackled research questions in the context of an online experiment. By conducting the study on the web, it was attempted to soften common laboratory setting limitations. As our research questions target RS performance, we did not use a standard digital library interface, where recommendations appear as a minor part of the webpage. We set up an interface that focused on user profile generation and performance evaluations instead. While this is surely a limitation when considering the applicability of the presented approaches to real world digital library scenarios, this setup had the advantage of not distracting the user from the actual study.

The LOD dataset of EconStor served as the backbone of our experiment. It was chosen, because it contains bibliographic information of an existing and frequently used digital library in the well-defined research area of economics. Additionally, many of its papers are annotated with SKOS concepts of the Thesaurus for Economics (STW). EconStor is among the largest Open Access servers in this field. Its current Linked Open Data dump comprises RDF triples of more than 108 k papers.[6] The online web experiment was designed to provide recommendations to participants for any EconStor input paper specified by him/her. Users were asked to only enter familiar publications that represent their research interests well. An AJAX-like interface supported participants in finding suitable publications, that were contained in our triple store, through an Apache Solr index (Fig. 5).[7] Once, a user had provided a publication and potentially additional constraints (Figs. 5 and 8) the engine was able to generate recommendations. However, user profiles would be created differently in a live application, e.g., based on session data or download statistics. But since previous preferences were not available to us, we simulated user interaction through the AJAX interface.

The online experiment was divided into four parts. In the first part (**Part I**), participants stated their gender, age, education and occupational status and provided some additional information regarding habits and preferences when searching for scientific literature. For each of the following parts (**Part II – Part IV**) participants were presented different recommendation lists for various query scenarios. Each one of the result lists contained up to six recommendations. Although, it would have probably been valuable to receive user evaluations for more recommendations, it was intended to keep the required user effort as small as possible in order to reduce the risk of study withdrawal. We applied strict within-subjects A/B tests. This was done because recent evidence has shown even small variations in the RS to cause comparably strong differences in performance evaluations [15]. These findings might be even more prevalent in the context of scientific publication retrieval. Research interests and literature preferences can be highly diverse among scientists. In order

[6] https://www.econstor.eu/about.

[7] http://lucene.apache.org/solr/.

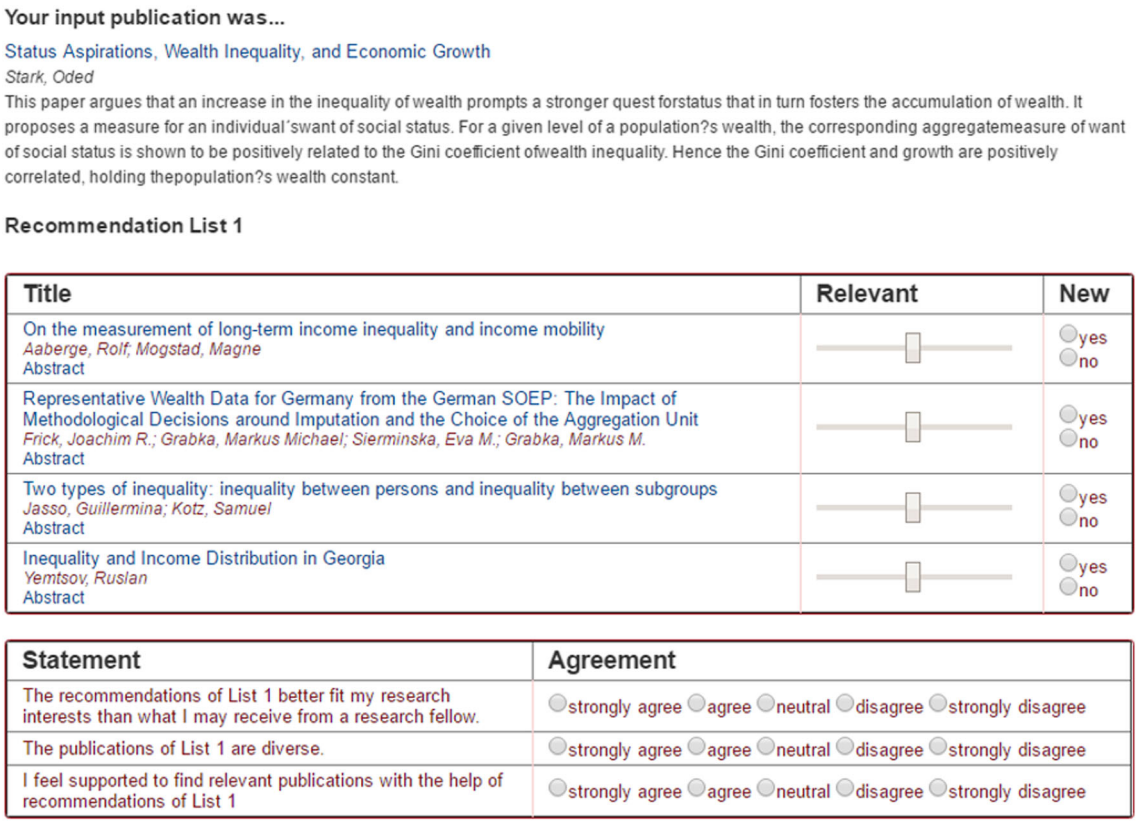

Fig. 6 Interface for evaluation of recommendation results (**Part II–IV**)

to avoid unintentional variations, we decided to let participants evaluate different recommendation approaches for the same input papers. Potential order effects, that might have been otherwise occurred, were controlled for by applying a subjects-by treatments design, which is commonly used in the social sciences [21]. In that, recommendation lists resulting from different approaches were displayed in randomized order. This procedure was followed through in **Part II** and **Part III**, where multiple recommendation approaches had to be directly compared on a single webpage. **Part II** was targeted at research question R_2 and asked subjects to both evaluate LOD-based recommendations as well as text-based recommendations. Text-based recommendations were derived from a standard "More Like This Query" from Apache Lucene/Solr based on the title, the author and the abstract of the paper.

For user evaluations, we decided to borrow methods from experiments in information retrieval and recommender systems. This is in line with our intention to utilize LOD recommendation queries as an effective extension of traditional IR in digital libraries. Subjects were instructed to use sliders to state if a publication was relevant to them (outer left side: lowest relevance, outer right side: highest relevance). Thus, the quality dimension of *accuracy* was addressed. Scores were handled as points on a 0 to 100 scale of a relevance slider for later analysis. Participants received a short snippet of the publication in order to allow for an informed decision. Relevance sliders and text snippets are commonly used in IR systems evaluations [16,17]. Additionally, for each of the recommended publications, participants were asked to state whether the respective item was known to him/her or not. Hence, evaluation of recommendation list *novelty* was ensured (Fig. 6). The *diversity* dimension was addressed through asking participants to give their agreement on a five point Likert-type question (*The publications of List ... are diverse*) targeted at the immediate user opinion.

To address research question R_1 and to increase general reliability, item-level evaluations were complemented with questions that are in line with frameworks for evaluation of a recommender system's general usefulness [9,10]. Two five point Likert items served to grasp user agreement to statements regarding general usefulness of the recommendations.

- Q_1. *The recommendations of List ... better fit my research interests than what I may receive from a research fellow*
- Q_2. *I feel supported to find relevant publications with the help of recommendations of List ...* .

Part III was concerned with R_3, where participants had to compare recommendations that resulted from the applica-

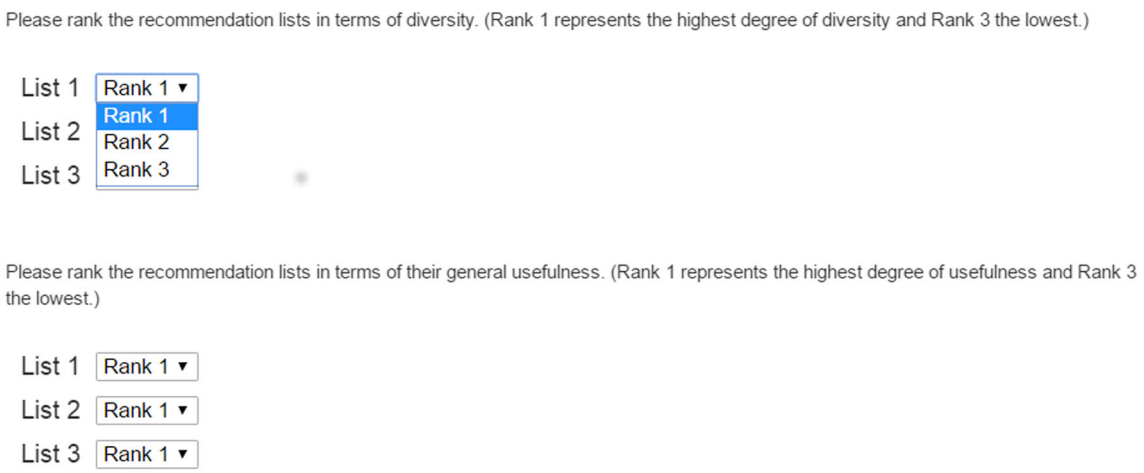

Fig. 7 Ranking recommendation lists according to usefulness and diversity (**Part III**)

tion of different granularity levels of inter-concept similarity ($sim_{1.0}$, $sim_{0.75}$ and $sim_{0.5}$). Inter-concept similarities were quantified with an adaptation of the measure proposed by Lin (see Eq. 15). It bases similarity calculation of two concepts c_1 and c_2 on the information content of their most specific common subsumer (c_0) [19]. Thereby, concept specificity was determined according to the number of descendants a concept subsumes [66].

$$sim(c_1, c_2) = \frac{IC(c_0)}{IC(c_1) + IC(c_2)} \tag{15}$$

The similarity metric by Lin was chosen because it is well suited to measure corpus-independent concept relatedness in a directed acyclic graph (DAG) [19]. Since the STW is organized as a polyhierarchic DAG-like structure, the measure can be seamlessly applied on the STW subgraph based on `skos:broader` and `skos:narrower` relations. Inter-concept similarities were precalculated for each concept pair of the STW Thesaurus with the help of the Semantic Measures Library [22]. They were integrated into the recommendation engine to be ready for use at runtime. After rating all items of each of the three recommendation lists individually, participants were asked to rank recommendations lists according to their diversity and general usefulness (see Fig. 7). In the last section (**Part IV**), we addressed research question R_4 by applying filters on results that were obtained from an intermediate level of inter-concept similarity ($sim_{0.75}$) as described in **Part III**. Participants were supported in choosing suitable constraints from an autocomplete interface (see Fig. 8).

When STW Keywords were chosen, an advanced query (see Sect. 3.5) was generated and issued against an RDF triple store. This form of basic graph pattern matching, as it is specified by the SPARQL syntax, was thereby used to retrieve all papers with annotations that are related to the specified user condition and similar to the input paper as well. The SPARQL query enabled users to apply all ancestors and related concepts of a specified keyword as a filter on a recommendation list.

Upon issuing a constraint-based recommendation query, participants were again shown an evaluation page, that was structured as in **Part II**. In addition to the already familiar user tasks (relevance sliders, novelty checkboxes and Likert-type questions), subjects were asked, if they agreed that the filter option had improved recommendation results (*Q4: The possibility to narrow the search with additional constraints helps me to get better recommendations.*).

A pretest preceded our online web experiment and helped to control for potential pitfalls (e.g., misleading instructions). The actual experiment was conducted in spring and summer 2016. A study link was posted in two online forums for economics in higher education. Additionally, randomly selected potential participants from different German economics research institutions were recruited via e-mail and phone calls to increase response rates. Subjects, that completed the online experiment (41 in total), were entered into a prize draw for a gift voucher. Two surveys were withdrawn from the evaluation due to incompleteness.

4.3 Evaluation measures

In concordance with our research questions and the experimental setup, participants evaluated all of the recommendation approaches. In this context, measures that are commonly applied in the areas of information retrieval and recommender systems were applied. Evaluations of all participants were aggregated for each approach based on the following measures.

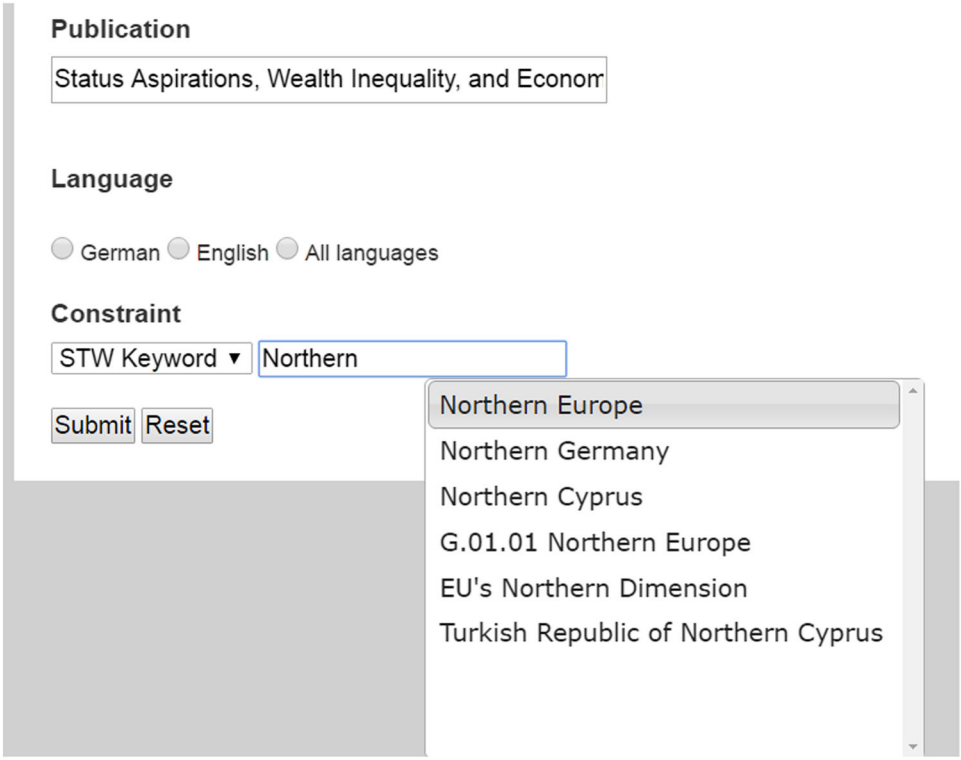

Fig. 8 Web interface for generating a user profile with constraints (**Part IV**)

Mean Relevance Score. Mean relevance scores from users for single items ($score_i$) were averaged for each recommendation list (Eq. 16).

$$mrs = \sum_{i=1}^{n} \frac{score_i}{n} \tag{16}$$

Mean Precision@k. For this measure, we assumed relevance scores to be binary values. Publications were either marked as "relevant" or as "irrelevant" [23]. In our experiment, we required a relevant item to have achieved a user score of 50 or higher on the relevance slider. The fraction of relevant publications ($|REL|$) as opposed to all publications ($|REC|$) that were ranked in the top k position of a user's recommendation list was determined for this purpose (Eq. 17).

$$prec@k = \frac{|REL|}{|REC|} \tag{17}$$

Normalized Cumulative Gain. In contrast to the above listed measures, NDCG does not capture relevance scores directly, but instead evaluates the ranking ability of an algorithm. Thereby, highly relevant documents, that turn up at the top of a recommendation list, achieve higher scores, than they would, when being listed at its bottom. NDCG also captures the value, that any user gains (dcg_n) (Eq. 18) from examining the presented ranking (G) as opposed to the gain he would receive when being confronted with a result list in ideal order ($idcg_n$) [24] (Eq. 19).

$$dcg_n = G[1] + \sum_{j=2}^{n} \frac{G[j]}{log_2\ j} \tag{18}$$

$$ndcg@n = \frac{dcg_n}{idcg_n} \tag{19}$$

Novelty. While accurate and familiar recommendations are useful because they build trust in a system [9], they might not be valuable in the long run due to their obviousness. This is especially true in the context of research paper recommender systems, where users typically seek to explore novel literature for their current topic of interest.

But even though it has been frequently argued to apply novelty measures [1,3,8–10] for recommender systems evaluation, few explicit novelty measures can be found in the literature. Castells [13] proposed to measure novelty by the log of the inverse popularity (i. e. user interaction frequency) of an item in a product catalogue. But this notion of popularity can be only applied on a dataset of reasonable size. Hence, we decided to follow the authors of [9] and [25], where novelty was defined as the ratio of the number of relevant documents not familiar to the user ($|NOV|$) and the total number of relevant documents ($|REC|$) for each user-specific recommendation list (Eq. 20). As it was done for precision, relevant publications were assumed to have achieved a relevance score

of 50 or higher. Novel items were identified by the respective user statements in the online experiment (see Sect. 4.2).

$$nv = \frac{|NOV|}{|REC|} \tag{20}$$

Diversity. In our experiment, diversity was measured from two different angles. On the one hand, user-based diversity ($divU$) was captured through the median of positions on the Likert scale (**Part I**, **Part III**) or the median of the reversed rank positions (**Part II**), where participants had to give their explicit feedback on diversity.

On the other hand, we determined implicit content-based diversity through evaluating item-to-item distances between each pair of the n items in a recommendation list (Eq. 22) as it was proposed by [13]. Thereby, we chose to obtain distance values of two research papers p_l and p_m with the help of the cosine similarity on publication features ($cos(p_l, p_m)$) (Eq. 21).

$$d(p_l, p_m) = 1 - cos(p_l, p_m) \tag{21}$$

$$divC = \frac{2}{n(n-1)} \sum_{l<m} d(p_l, p_m) \tag{22}$$

5 Results

The online experiment had an average processing time of approx. 17 mins. 36 participants were between 20–40 years old and most of the study subjects (23 in total) were male. The vast majority of the sample either studied at undergraduate level (4 subjects) or were employed as PhD or postdoc fellows (33 subjects). Almost all participants (37) listed Google Scholar as one of their primary sources for publication retrieval. Participants mentioned Open Access databases (26) and library catalogues (21) as well. Fewer participants mentioned commercial databases (13) or other sources (2). The majority of the subjects (26) stated that literature search in their research area is "easy" or "very easy", while the remaining participants found it to be at least "manageable".

R_1. Due to the still acceptable reliability of the Likert-type items addressing general usefulness (Cronbach's $\alpha = 0.69$), we performed a median-based aggregation and tested its correlations with accuracy, novelty and diversity. Table 1 depicts results and p values accordingly. It can be seen that only accuracy values seem to be correlated with usefulness at acceptable levels of significance. Despite there was no strong significant connection between novelty and general usefulness in the data, when looking at participants responses regarding RS performance as compared to suggestions from research fellows ($Q1$), novel recommendations showed to be slightly contributing to users opinions on usefulness, though being only marginally significant ($\rho = 0.208$, $p < 0.1$).

Table 1 Spearman's ρ for recommendation quality (R_1)

	Usefulness	
	ρ	p
Acc. ($mrs/100$)	.481	.000
Acc. ($ndcg@n$)	.184	.106
Acc. ($prec@n$)	.396	.000
Nov. (nv)	.156	.209
Div. ($divU$)	−.007	.955
Div. ($divC$)	−.009	.940

Even though, these results are in line with previous research, that confirmed trust-building and relevant recommendations to be the most important factor [1], they seem quite surprising in the digital library context, where one might expect users to strongly favor novel recommendation results. In addition to that, diversity seems to be inferior for our sample, as opposed to findings from other authors [1,6]. But the low to zero impact of diversity might be attributable to the fact, that in our study users only had to evaluate rather short recommendation lists.

R_2. Participants quality evaluations of LOD recommendations as well as text-based recommendations were examined. Wilcoxon signed rank pairwise comparisons were made for assessments of general usefulness and—due to non-normality—also for the novelty and diversity values. Text-based recommendations were slightly better in terms of general usefulness ($V = 280.5$, $p < 0.050$). They showed increased implicit diversity scores as well ($divC$) ($V = 622$, $p < 0.001$), but this might not affect recommendation quality at all, as the previous section indicated.

Systematic differences for other aspects could not be verified. The two recommendation methods do not seem to outperform the other with regard to novelty (nv), explicit diversity ($divU$) or accuracy (mrs, $ndcg@n$, $prec@n$). Based on these diverging findings and the seemingly small differences in means and deviation among the different evaluation measures, it can be assumed, that text-based recommendations might be only slightly better than LOD recommendations (see Table 2).

R_3. We compared evaluations for recommendations that were obtained from 3 granularity levels of inter-concept similarity ($sim_t, t \in \{1.0, 0.75, 0.5\}$). Non-parametric repeated measures Friedman tests were conducted to investigate significance in differences of means. Participants' rankings of the 3 approaches did not show to be significantly different for the three approaches, neither with regard to general usefulness ($\chi^2 = 4.015$, $p = 0.134$), nor with regard to diversity ($\chi^2 = 2.882$, $p = 0.237$). In concordance with

Table 2 Descriptive statistics for LOD & Text Rec. (R_2)

	LOD Rec.		Text Rec.	
	M	SD	M	SD
Acc. ($mrs/100$)	.509	.216	.503	.233
Acc. ($ndcg@n$)	.585	.316	.567	.283
Acc. ($prec@n$)	.685	.412	.542	.361
Nov. (nv)	.888	.110	.913	.115
Div. ($divC$)	.640	.148	.747	.179

Table 3 Descriptive statistics for different inter-concept similarities (t) (R_3)

	$t = 1.0$		$t = 0.75$		$t = 0.5$	
	M	SD	M	SD	M	SD
Acc. ($mrs/100$)	.490	.215	.426	.230	.476	.224
Acc. ($ndcg@n$)	.863	.108	.817	.176	.847	.118
Acc. ($prec@n$)	.568	.292	.467	.350	.512	.314
Nov. (nv)	.623	.622	.777	.122	.763	.157
Div. ($divC$)	.666	.566	.634	.400	.715	.691

that, a Friedman test of novelty ($\chi^2 = 1.263$, $p = 0.532$) and an additional repeated measures ANOVA of accuracy values (mrs) could not verify any systematic differences ($F(1, 38) = 0.308$, $p = 0.582$).

These are interesting findings considering the fact that each similarity level also produced divergent recommendations. This is indicated by the mean of Jaccard similarity coefficients (*inter-list similarity*) among recommendation lists being considerably low ($j = 0.214$). A subsequent Friedman test on concept-based diversity scores ($divC$) confirmed considerable differences within the lists ($\chi^2 = 1.960$, $p = 0.375$, $p < 5e-07$). These *intra-list similarity* scores were significantly lower with higher diversity scores, when the flexible similarity detection method was applied, as was revealed by Wilcoxon post-hoc tests. ($H_1 : \mu_{0.75} > \mu_{1.0}, V = 170, p < 5e-06$; $H_1 : \mu_{0.5} > \mu_{1.0}, V = 694, p < 5e-05$).

Based on these results, one can assume, that our approach might have a diversifying effect, that considerably changes ranking results, which could be of value in exploratory search tasks. Additionally, bearing in mind the inter-result set varieties on the one hand and the lack of systematic differences of user evaluations on the other hand, it is hypothesized that an exact concept-to-concept matching ($t = 1.0$) will not automatically lead to the best results. Hence, it would be desirable to enable users to flexibly adjust similarity calculation to different levels of granularity in order to obtain recommendations that best fit their needs.

But this can only be the case if the perceived quality of generated recommendations is sufficiently high. Table 3 shows that automatic suggestions achieved acceptable to good results for all three values of t. This led us to the conclusion that the approach of *flexible similarity detection* should be further explored for exploratory search. Hence, we simulated browsing behavior by selecting evaluations of the best performing approach, as indicated by the highest ranked list in terms of general usefulness for each subject and compared it to the evaluations obtained through exact concept-to-concept matchings. A paired t-test showed mrs scores to be marginally significantly better, when browsing is enabled ($t_{38} = 1.315$, $p < 0.1$). Additional Wilcoxon signed ranked pairwise comparisons resulted in the same finding for $prec@n$ ($V = 68.5$, $p < 0.06$) and content-based diversity $divC$ ($V = 152$, $p < 0.05$). Other significant differences, e.g., in the areas of novelty could not be identified.

R_4. The fourth research question seeked to analyze how users perceive recommendations that can be adapted to their individual preferences. We compared the performance of the recommendation algorithm with an intermediate level of inter-concept similarity ($sim_{0.75}$) (**Part III**) to results from **Part IV** that were obtained from the same algorithm with additional conditions ($sim_{0.75}$ + filter). In order to stick closely to the original within-subjects design of our experimental setup, we made sure, that users chose the same input paper in each **Part III** and **IV**. Results from subjects with non-identical input papers were removed from data analysis.

In our experiment, recommendations that resulted from previous user filtering achieved better results in terms of accuracy values than those that were obtained without additional constraints (see Fig. 9). A t-test on mrs values ($t_{26} = 3.141$) and subsequent Wilcoxon ranked pairwise comparisons for $prec@n$ ($V = 167$, $p = .011$) and $ndcg@n$ ($V = 312$, $p = .001$) scores proved to be significant for this hypothesis. In addition to that, when asked directly about their attitude, many participants either agreed to the statement (46.4%), that filtering improves recommendation results (Q_4) or at least stated their neutrality (17.9%) toward it. Other differences in performance either with regard to novelty or with regard to diversity did not prove to be significant.

Apart from user evaluations, differences in recall were investigated as well. For this purpose, we took a closer look at results set sizes when users opted to filter their recommendation lists with keywords ($N = 15$). When advanced graph pattern matching was applied (see SPARQL query in Sect. 3), only 1 participant received an empty result set as opposed to 3 participants that would have not received results, when exact keyword matching was implemented. But given the small sample set size, final conclusions should only be drawn after additional research.

In summary, the web-based user experiment gave first insights into the research question of how LOD informa-

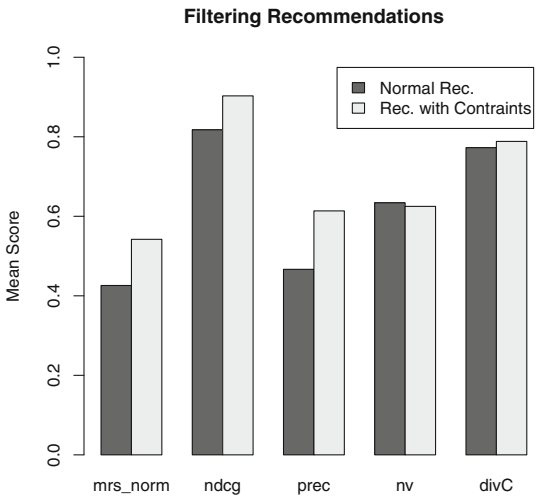

Fig. 9 Evaluation results for recommendations with and without constraints (R_4)

tion can potentially leverage information retrieval in the digital library context. It revealed tendencies in terms of recommendation quality for users, who were confronted with novel LOD-enabled retrieval strategies. Although the findings indicate that users might profit from the presented approaches, conclusions have to be drawn cautiously given the limited scope of the experiment. The results will have to be backed by further studies in other application domains and with a bigger sample of participants in order to be fully able to claim superiority of the presented approaches.

6 Conclusion

In this paper we have investigated, how bibliographic datasets from LOD endpoints can be utilized for recommender systems in digital libraries. In this context, key application areas were identified. Publicly available datasets from libraries and other information providers could potentially be of high value, in case data in local repositories is limited or insufficient. They provide rich content, of which Linked Data-enabled Knowledge Organization Systems (SKOS vocabularies) might offer powerful means of resource access.

We conducted a web-based experiment to test the feasibility of LOD recommendations in the digital library context. Users were asked to give their opinions on recommendations that were generated with SKOS annotations as opposed to text-based recommendations. The results showed that LOD recommendations perform slightly poorer than text-based recommendations. But this tendency seems to be rather weak, as is indicated by distributions of evaluation measures for the two methods. Based on these findings, it is hypothesized that RDF data can be a viable alternative, when no fulltext is available.

Another advantage of LOD recommendations is that they can be computed on-the-fly. Thus, automatic suggestions are generated through an API-like interface without imposing data maintenance or NLP preprocessing workload on the side of the digital library provider.

In addition to feasibility evaluation, we proposed two novel content-based recommendation approaches, namely *flexible similarity detection* and *contraint-based recommendations*, which might support users in finding suitable publications when other query strategies have been exhausted.

As such, *flexible similarity detection* facilitates browsing and exploratory searching. It enables users to refine a recommendation request based on different levels of inter-concept similarities until results best fit their needs. In the area of *constraint-based recommendations*, similarity calculation is combined with graph pattern matching, that is known from common SPARQL queries. LOD repositories can be accessed with query requests through the application of filters. Inexperienced users could profit from this approach, because they might just be able to formulate straightforward preferences (e.g., the subject of an article) and can rely on their user profiles or session histories to reflect their interests well enough to serve as a proxy for an advanced query. Expert users, on the other hand, might find an added value in formulating an expressive filter condition for a long list of recommendations in order to fully explore the semantic space to meet their demands.

We conducted a web experiment, where both strategies were evaluated. The study results indicate that *flexible similarity detection* as well as *contraint-based recommendations* can be promising alternatives for retrieval access in the digital library context. Our web application serves as a proof-of-concept implementation, as we seeked to provide a ready-to-use system during the experiment. We showed that users can be assisted in formulating filter conditions for recommendation results through an auto-complete list. Complementary and suitable user interface features, e.g., for browsing based on *flexible similarity detection* will have to be developed and validated in the future. In addition, subsequent and more comprehensive investigations will have to show if the initial evaluation results can be replicated.

In summary, LOD recommendations offer a powerful alternative for retrieval access in the digital library context through complementing local resources and through the possibility for advanced query strategies both for inexperienced and expert users.

Acknowledgements We would like to thank reviewers for their helpful suggestions. Special thanks goes to the Leibniz Information Centre for

Economics (ZBW) for providing data (fulltexts and STW descriptors as LOD dataset), consulting and testing with respect to the overall application scenario. Additionally, we would like to thank Mandy Sachse for helping with the recruitment of participants.

References

1. Pu, P., Chen, L., Hu, R.: Evaluating recommender systems from the user's perspective: survey of the state of the art. User. Model. User-Adap. Inter. **22**, 317–355 (2012). doi:10.1007/s11257-011-9115-7
2. Lam, S.K., Frankowski, D., Riedl, J.: Do you trust your recommendations? An exploration of security and privacy issues in recommender systems. In: Müller, G. (ed.) Emerging trends in information and communication security, pp. 14–29. Springer, Berlin (2006). doi:10.1007/11766155_2
3. Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.T.: Evaluating collaborative filtering recommender systems. ACM Trans. Inf. Syst. (TOIS) **22**, 5–53 (2004). doi:10.1145/963770.963772
4. Adamopoulos, P., Tuzhilin, A.: On unexpectedness in recommender systems: or how to expect the unexpected. ACM Trans. Intell. Syst. Technol. (TIST) (2011). doi:10.1145/2559952
5. Ge, M., Delgado-Battenfeld, C., Jannach, D.: Beyond accuracy: evaluating recommender systems by coverage and serendipity. In: Proceedings of the 4th ACM Conference on Recommender Systems (RecSys '10), pp. 257–260 (2010). doi:10.1145/1864708.1864761
6. Ziegler, C.N., McNee, S.M., Konstan, J.A., Lausen, G.: Improving recommendation lists through topic diversification. In: Proceedings of the 14th International Conference on World Wide Web (WWW '05), pp. 22–32 (2005). doi:10.1145/1060745.1060754
7. Adomavicius, G., Kwon, Y.: Toward more diverse recommendations: item re-ranking methods for recommender systems. In: Proceedings of the 19th Workshop on Information Technology and Systems (WITS '09) (2009)
8. McNee, S.M., Riedl, J., Konstan, J.A. (2006) Being accurate is not enough : how accuracy metrics have hurt recommender systems. Ext. Abs. Human Factors Comput. Syst. doi:10.1145/1125451.1125659
9. Swearingen, K., Sinha, R.: Beyond algorithms : an HCI perspective on recommender systems. In: Proceedings of the ACM SIGIR 2001 Workshop on Recommender Systems (2001)
10. Pu, P., Chen, L., Hu, R.: A user-centric evaluation framework for recommender systems. In: Proceedings of the 5th ACM Conference on Recommender Systems (RecSys '11), pp. 157–164. doi:10.1145/2043932.2043962 (2011)
11. Cosley, D., Lam, S.K., Albert, I., Konstan, J.A., Riedl, J.: Is seeing believing?: how recommender system interfaces affect users' opinions. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 585–592 (2003). doi:10.1145/642611.642713
12. Beel, J., Gipp, B., Langer, S.: Research-paper recommender systems: a literature survey. Int. J. Digit. Libr. **17**, 305–338 (2016). doi:10.1007/s00799-015-0156-0
13. Castells, P., Vargas, S., Wang, J.: Novelty and diversity metrics for recommender systems: choice, discovery and relevance. In: International Workshop on Diversity in Document Retrieval (DDR 2011) at the 33rd European Conference on Information Retrieval (ECIR 2011) (2011)
14. Beel, J., Breitinger, C., Langer, S., Lommatzsch, A., Gipp, B.: Towards reproducibility in recommender-systems research. User. Model. User-Adap. Inter. **26**, 69–101 (2016). doi:10.1007/s11257-016-9174-x
15. Hersh, W., Turpin, A., Price, S., Chan, B., Kramer, D., Sacherek, L., Olson, D.: Do batch and user evaluations give the same results? In: Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '00), pp. 17–24 (2000). doi:10.1145/345508.345539
16. Manning, C.D., Raghavan, P., Schütze, H.: Introduction to information retrieval. Cambridge University Press, Cambridge (2008)
17. Ruthven, I., Lalmas, M., Van Rijsbergen, K.: Incorporating user search behavior into relevance feedback. J. Am. Soc. Inf. Sci. Technol. **54**, 529–549 (2003)
18. Rui, Y., Huang, T.S.: A novel relevance feedback technique in image retrieval. In: Proceedings of the 7th ACM International Conference on Multimedia (Multimedia '99), pp. 67–70 (1999). doi:10.1145/319878.319896
19. Lin, D.: An information-theoretic definition of similarity. In: Proceedings of the 15th International Conference on Machine Learning (ICML 1998), pp. 296–304 (1998)
20. Neubert, J.: Bringing the 'thesaurus for economics' on to the web of linked data. In: Proceedings of the Linked Data on the Web Workshop (LDOW '2009) (2009)
21. Kirk, R.E.: Experimental design-procedures for the behavioral sciences. Sage, Los Angeles (2013)
22. Harispe, S., Ranwez, S., Janaqi, S., Montmain, J.: The semantic measures library and toolkit: fast computation of semantic similarity and relatedness using biomedical ontologies. Bioinform **30**, 740–742 (2014)
23. Agichtein, E., Brill, E., Dumais, S.: Improving web search ranking by incorporating user behavior information. In: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '06), pp. 19–26 (2006). doi:10.1145/1148170.1148177
24. Järvelin, K., Kekäläinen, J.: Cumulated gain-based evaluation of IR techniques. ACM Trans. Inf. Syst. (TOIS) **20**, 422–446 (2002). doi:10.1145/582415.582418
25. El Demerdash, O., Bergler, S., Kosseim, L., Langshaw, P.K.: Developing AMIE: an adaptive multimedia integrated environment. In: International Workshop on Adaptive Multimedia Retrieval, pp. 65–78 (2005). doi:10.1007/11670834_6
26. Bizer, C., Heath, T., Berners-Lee, T.: Linked data-the story so far. Semant. Serv., Interoper. Web Appli.: Emer. Concepts, pp. 205–227 (2009)
27. Mönnich, M., Spiering, M.: Adding value to the library catalog by implementing a recommendation system. D-Lib Magazine 14 (2008)
28. Sarwar, B., Karypis, G., Konstan, J., Riedl, J.: Item-based collaborative filtering recommendation algorithms. In: Proceedings of the 10th International Conference on World Wide Web (WWW '01), pp. 285–295 (2001). doi:10.1145/371920.372071
29. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Trans. Knowl. Data Eng. **17**, 734–749 (2005). doi:10.1109/TKDE.2005.99
30. Yang, C., Wei, B., Wu, J., Zhang, Y., Zhang, L.: CARES: a ranking-oriented CADAL recommender system. In: Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '09), pp. 203–212 (2009). doi:10.1145/1555400.1555432
31. Di Noia, T., Mirizzi, R., Ostuni, V.C., Romito, D., Zanker, M.: Linked open data to support content-based recommender systems. In: Proceedings of the 8th International Conference on Semantic Systems (I-SEMANTICS '12) (2012). doi:10.1145/2362499.2362501
32. Wenige, L., Ruhland, J.: Flexible on-the-fly recommendations from linked open data repositories. In: Proceedings of the International Conference on Business Information Systems (BIS 2016), pp. 43–54 (2016). doi:10.1007/978-3-319-39426-8_4
33. Carpineto, C., Romano, G.: A survey of automatic query expansion in information retrieval. ACM Comput. Surv. (CSUR) (2012). doi:10.1145/2071389.2071390

34. Hajra, A., Latif, A., Tochtermann, K.: Retrieving and ranking scientific publications from linked open data repositories. In: Proceedings of the 14th International Conference on Knowledge Technologies and Data-driven Business (i-KNOW '14) (2014). doi:10.1145/2637748.2638436
35. Shekarpour, S., Höffner, K., Lehmann, J., Auer, S.: Keyword query expansion on linked data using linguistic and semantic features. In: Proceedings of the IEEE 7th International Conference on Semantic Computing (ICSC), pp. 191–197 (2013). doi:10.1109/ICSC.2013.41
36. Di Noia, T., Mirizzi, R., Ostuni, V.C., Romito, D.: Exploiting the web of data in model-based recommender systems. In: Proceedings of the 6th ACM Conference on Recommender Systems, pp. 253–256 (2012). doi:10.1145/2365952.2366007
37. Ruotsalo, T., Haav, K., Stoyanov, A., Roche, S., Fani, E., Deliai, R., Mäkelä, E., Kauppinen, T., Hyvönen, E.: SMARTMUSEUM: a mobile recommender system for the web of data. Web Semant.: Sci., Serv. Agents World Wide Web **20**, 50–67 (2013)
38. Wu, Z., Palmer, M.: Verbs semantics and lexical selection. In: Proceedings of the 32nd Annual Meeting on Association for Computational Linguistics (ACL '94), pp. 133–138 (1994). doi:10.3115/981732.981751
39. Resnik, P.: Using information content to evaluate semantic similarity in a taxonomy. Accessed 21 January 2017 arXiv preprint arXiv:cmp-lg/9511007. (1995)
40. Kules, B., Capra, R.: Influence of training and stage of search on gaze behavior in a library catalog faceted search interface. J. Am. Soc. Inf. Sci. Technol. **63**, 114–138 (2012). doi:10.1002/asi.21647
41. Wakeling, S., Clough, P., Sen, B.: Investigating the potential impact of non-personalized recommendations in the OPAC: Amazon vs. WorldCat.org. In: Proceedings of the 5th Information Interaction in Context Symposium (IIiX '14), pp. 96–105 (2014). doi:10.1145/2637002.2637015
42. Ballard, T., Blaine, A.: User search-limiting behavior in online catalogs: comparing classic catalog use to search behavior in next-generation catalogs. New Lib. World **112**, 261–273 (2011). doi:10.1108/03074801111136293
43. Kules, B., Capra, R., Banta, M., Sierra, T.: What do exploratory searchers look at in a faceted search interface? In: Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '09), pp. 313–322 (2009). doi:10.1145/1555400.1555452
44. Antell, K., Huang, J.: Subject searching success: transaction logs, patron perceptions, and implications for library instruction. Ref. Us. Serv. Q. **48**, 68–76 (2008)
45. Yang, S.Q., Wagner, K.: Evaluating and comparing discovery tools: how close are we towards next generation catalog? Lib. Hi Tech **28**, 690–709 (2010)
46. Konstan, J.A., Kapoor, N., McNee, S.M., Butler, J.T.: Techlens: Exploring the use of recommenders to support users of digital libraries. CNI Fall Task Force Meeting Project Briefing. Accessed 21 January 2017 (2005)
47. Middleton, S.E., Shadbolt, N.R., De Roure, D.C.: Ontological user profiling in recommender systems. ACM Trans. Inf. Syst. **22**, 54–88 (2004). doi:10.1145/963770.963773
48. Porcel, C., Moreno, J.M., Herrera-Viedma, E.: A multi-disciplinar recommender system to advice research resources in university digital libraries. Exp. Syst. Appli. **36**, 12520–12528 (2009). doi:10.1016/j.eswa.2009.04.038
49. Vellino, A.: Recommending research articles using citation data. Lib. Hi Tech **33**, 597–609 (2015). doi:10.1108/LHT-06-2015-0063
50. Webster, J., Jung, S., Herlocker, J.: Collaborative filtering: a new approach to searching digital libraries. New Rev. Inform. Netw. **10**, 177–191 (2004). doi:10.1080/13614570500054229
51. Kabutoya, Y., Sumi, R., Iwata, T., Uchiyama, T., Uchiyama, T.A.: Topic model for recommending movies via linked open data. In: Proceedings of the IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology (WI-IAT), pp. 625–630 (2012). doi:10.1109/WI-IAT.2012.23
52. Ostuni, V.C., Di Noia, T., Mirizzi, R., Romito, D., Di Sciascio, E.: Cinemappy: a context-aware mobile app for movie recommendations boosted by DBpedia. In: Proceedings of the 2012 International Conference on Semantic Technologies Meet Recommender Systems & Big Data (SeRSy '12), pp. 37–48 (2012)
53. Passant, A.: dbrec - music recommendations using dbpedia. In: Proceedings of the 9th International Semantic Web Conference, (ISWC 2010), pp. 209–224 (2010). doi:10.1007/978-3-642-17749-1_14
54. Peska, L., Vojtas, P. (2013) Enhancing recommender system with linked open data. Flex. Query Answ. Syst. doi:10.1007/978-3-642-40769-7_42
55. Celma, Ò., Serra, X.: FOAFing the music: bridging the semantic gap in music recommendation. In: Proceedings of the 5th International Semantic Web Conference (ISWC 2006), pp. 927-934 (2008). doi:10.1007/11926078_67
56. Mobasher, B., Jin, X., Zhou, Y.: Semantically enhanced collaborative filtering on the web. Web Mining: From Web to Semantic Web, pp. 57–76 (2004). doi:10.1007/978-3-540-30123-3_4
57. Ostuni, V.C., Di Noia, T., Di Sciascio, E., Mirizzi, R.: Top-n recommendations from implicit feedback leveraging linked open data. In: Proceedings of the 7th ACM Conference on Recommender Systems (RecSys '13), pp. 85–92 (2013). doi:10.1145/2507157.2507172
58. Heitmann, B., Hayes, C.: Using linked data to build open, collaborative recommender systems. AAAI Spring Symp.: Link. Data Meets Artif. Intelli., pp. 76–81 (2010)
59. Ziegler, C.N.: Semantic web recommender systems. Current Trends Data. Technology-EDBT 2004 Workshops, pp. 78–89 (2005). doi:10.1007/978-3-540-30192-9_8
60. Summers, E., Isaac, A., Redding, C., Krech, D.: LCSH, SKOS and linked data. In: Proceedings of the International Conference on Dublin Core and Metadata Applications, pp. 25–33 (2008)
61. Tuominen, J., Frosterus, M., Viljanen, K., Hyvönen, E.: ONKI SKOS server for publishing and utilizing SKOS vocabularies and ontologies as services. In: Proceedings of the 6th European Semantic Web Conference (ESWC 2009), pp. 768–780 (2009). doi:10.1007/978-3-642-02121-3_56
62. Neubert, J., Tochtermann, K.: Linked library data: offering a backbone for the semantic web. In: Lukose D, Ahmad AR, Suliman A (ed) Knowledge Technology, Springer, Berlin pp. 37–45. doi:10.1007/978-3-642-32826-8_4
63. Caracciolo, C., Stellato, A., Rajbahndari, S., Morshed, A., Johannsen, G., Jaques, Y., Keizer, J.: Thesaurus maintenance, alignment and publication as linked data: the AGROVOC use case. Inter. J. Metadata, Seman. Ontolog. **7**(1), 65–75 (2012). doi:10.1504/IJMSO.2012.048511
64. Liao, I., Hsu, W., Cheng, M., Chen, L.: A library recommender system based on a personal ontology model and collaborative filtering technique for English collections. Electron. Lib. **28**, 386–400 (2010). doi:10.1108/02640471011051972
65. Mooney, R.J., Roy, L.: Content-based book recommending using learning for text categorization. In: Proceedings of the 5th ACM Conference on Digital Libraries (DL '00), pp. 195–204 (2004). doi:10.1145/336597.336662
66. Seco, N., Veale, T., Hayes, J.: An intrinsic information content metric for semantic similarity in WordNet. In: Proceedings of the 16th European Conference on Artificial Intelligence (ECAI '04) (2004)
67. Zapilko, B., Schaible, J., Mayr, P., Mathiak, B.: TheSoz: A SKOS representation of the thesaurus for the social sciences. Semant. Web **4**(3), 257–263 (2013)
68. Mayr, P., Petras, V.: Cross-concordances: terminology mapping and its effectiveness for information retrieval. In: Proceedings of the 74th IFLA World Library and Information Congress (2008)