Linked Data Metrics for Flexible Expert Search on the Open Web

Milan Stankovic[1,2], Jelena Jovanovic[3], and Philippe Laublet[2]

[1] hypios, 187 rue du Temple, 75003 Paris, France
[2] STIH, Université Paris-Sorbonne, 28 rue Serpente, 75006 Paris, France
[3] Université de Belgrade, Jove Ilica 154, 11000 Belgrade, Serbia
`milan.stankovic@hypios.com, jeljov@fon.rs,`
`philippe.laublet@paris-sorbonne.fr`

Abstract. As more and more user traces become available as Linked Data Web, using those traces for expert finding becomes an interesting challenge, especially for the open innovation platforms. The existing expert search approaches are mostly limited to one corpus and one particular type of trace – sometimes even to a particular domain. We argue that different expert communities use different communication channels as their primary mean for communicating and disseminating knowledge, and thus different types of traces would be relevant for finding experts on different topics. We propose an approach for adapting the expert search process (choosing the right type of trace and the right expertise hypothesis) to the given topic of expertise, by relying on Linked Data metrics. In a gold standard-based experiment, we have shown that there is a significant positive correlation between the values of our metrics and the precision and recall of expert search. We also present hy.SemEx, a system that uses our Linked Data metrics to recommend the expert search approach to serve for finding experts in an open innovation scenario at hypios. The evaluation of the users' satisfaction with the system's recommendations is presented as well.

Keywords: Expert Finding, Linked Data, Linked Data Metrics, Expertise Hypothesis.

1 Introduction

Developing innovation before competitors has been the key to survival and success in many industries. Standard approaches to this include having an in-house R&D department, hiring consultants, hiring experts from expert listing websites etc. Faced with long time that those approaches take to pass from the formulation of an innovation problem to its final solution, the companies turn to open innovation [1] platforms such as hypios.com, NineSigma.com and Innocentive.com, where they post calls for solutions to their problems. Such an approach is intended to bring more diversity to the considered solutions than closed, domain-focused expert communities, and other legacy approaches could offer. Experts from various (sometimes unforeseen) domains propose solutions and the innovation seeker picks the best solutions to buy and implement.

Our practice at hypios has shown that simply posting a problem online is not enough to attract diverse and innovative solutions; experts need to be invited to participate and the innovation challenge has to be introduced to them in a relevant manner. Open innovation platforms are thus faced with the challenge of identifying potential solvers – a challenge similar to but still different enough from the traditional expert search. Experts in the sense of open innovation are people who are capable of bringing a solution to a particular problem, sometimes by considering the problem from an unexpected perspective. In addition, they are not necessarily the best ranked experts for the problem topic.

Conveniently enough for the expert identification task, Web users produce more and more content, and more and more information about them is available on the Web. Those traces about users could serve as a ground for expert identification, and in fact a great number of expert search approaches has been proposed to identify and rank experts based on user traces. For instance, Buitelaar & Eigner, [2] rely on research papers, Kolari et al. [3] and Chua [4] on users' blogs, whereas Demartini [5] uses Wikipedia pages to infer and rank expertise. Content generation is not the only relevant activity for identifying experts; traces about users' bookmarking have also been used for expert finding [6], as well as question answering [7], and obtaining research grants [8].

However, the expertise search approaches found in literature often limit themselves to one or a few types of user traces and optimize their expert identification and ranking algorithms for the chosen type(s) of trace. This makes them and the systems built on top of them inflexible, in terms that they could not be easily modified or extended to accommodate a new type of user trace and/or a new kind of expertise hypothesis. In addition, their corpuses usually concern a particular domain of expertise and, accordingly, a particular expert community. As different expert communities tend to use different channels as their primary means of communication and dissemination of results, the types of representative user traces would be different for different topic of expertise. Since broad and extensible domain coverage is one of the core design principles for a Web-wide expert search system for open innovation problem, there is a strong need for versatility of such system in terms of the user traces to be used for expert identification. The state-of-the-art approaches lack this versatility as they are most often bound to a particular type of user trace due to the necessity to construct a specific data extraction approach to harvest the data from the chosen type of source.

The growing availability of user traces in structured form within the Linked Open Data (LOD) Cloud (e.g., DBLP, Semantic ChrunchBase, Twarql, RDFohloh[1] etc.) offers huge opportunities for overcoming the identified limitations of the existing expertise search approaches, as we aim to demonstrate in this paper. In particular, we propose a flexible solution for expert search in the context of open innovation problems, by leveraging the LOD Cloud and Linked Data metrics. The proposed solution helps users to select the most suitable way for identifying experts for a given topic of expertise, and subsequently applies the users' selection to the structured data about user traces on the Web (i.e., LOD) to retrieve the list of experts.

---

[1] Exact URIs for those data sets containing publications, company profiles, tweets, and project descriptions respectively, can be found on the official clickable LOD graph
`http://richard.cyganiak.de/2007/10/lod/imagemap.html`

The rest of the paper is structured as follows: in Section 2 we formalize the notions of user traces and expertise hypotheses – the defining traits of an expert search approach. In Section 3, we introduce Linked Data metrics, and demonstrate how they can help to choose an appropriate expert search approach. Section 4 presents hy.SemEx, our proof of the concept system for expertise hypothesis recommendation and expert search on LOD. In Section 5 we present related work and conclude the paper in Section 6.

2   User Traces and Expertise Hypotheses

In this section we define the basic notions relevant to the task of finding experts by using linked data on the Web: the notion of *user trace* that represents some form of evidence that a particular user might be an expert, and the notion of *expertise hypothesis* that defines the rule by which one concludes that a particular user is an expert based on the available user traces.

2.2   The Model of User Traces on the Linked Data Web

Definition 1. *A user trace in the context of our work is an information object that can be found on the Web; it concerns a particular Web user, and is about a particular topic of expertise. The trace is a form of evidence of the expertise that the user concerned by the trace possesses in the topic that the trace is about.*

For instance, the capability of a user to write a blog post on a particular topic may be considered as an evidence of his expertise in this topic. The trace in this case is the blog post in question.

On LOD, traces are presented as <*user, trace, topic*> triples, where an information object (i.e., a trace) is related to a user (e.g., a foaf:Agent[2]) and a topic. The former relation could be the relation of creation, possession, participation (in a professional event that the trace is about), or a mention of the user in the trace, etc. The latter relation is usually formalized through the dct:subject relation pointing to a concept of general knowledge coming from a Web-based knowledge base, such as Freebase.org or DBPedia.org.

2.3   Expertise Hypothesis

Definition 2. *Expertise hypothesis is an inference mechanism, which defines how to use the information contained in a user trace to identify and/or rank experts.*

For instance, an expertise hypothesis might be: if a user has written at least 3 blog posts on topic x, he may be considered as expert on topic x. Hypotheses may be aimed at the identification of experts (like the one we just gave) or they could serve to rank experts among each other. Each legacy expert search approach relies on one or a small set of implicit expertise hypotheses. We have provided a detailed overview of expertise hypotheses from the expert search literature in [9].

---

[2] All the prefixes used in this paper can be dereferenced on `http://prefix.cc`

Apart from the identifying vs. ranking nature of hypotheses, they are also characterized by the type of trace they use, the level of indirection between the user and the topic of expertise, the restrictiveness, etc. In order to capture the metadata about hypotheses and allow for their meaningful exchange between expert search systems, we have created the Expertise Hypothesis Ontology (EHO)[3]. An additional important use of EHO is in describing the provenance of expert search results produced by expert finding systems. Specifically, those systems could use EHO to state which hypotheses they used to obtain the results and on which dataset. This further allows for determining whether and to what extent the results of different expert search systems could be compared and integrated. Code 1 exemplifies a hypothesis' metadata represented using EHO. A basic description includes at least a human-readable description (eho:hasTextualStatement), a rule format (eho:hasFormat) and a URL where the executable hypothesis' rule is stored (eho:hasRuleURL). In our case the rule is a SPARQL query that can retrieve experts on a given topic. The topic of choice is passed as a parameter of the query (eho:hasParameter).

```
@prefix swrc: http://swrc.ontoware.org/ontology-07#
@prefix eho: http://ontologies.hypios.com/eho#
@prefix swc: http://data.semanticweb.org/ns/swc/ontology#
@prefix dct: http://purl.org/dc/terms/

:h1 a eho:ExpertSelectionHypothesis ;
      eho:hasTextualStatement "If a user wrote a scientific publication on topic X than he
is an expert on topic X" ;
      eho:hasTraceType swrc:Publication, swrc:InProceedings, swc:Paper ;
      eho:hasRule :r1_1 .
:r1_1 a eho:ExpertiseHypothesisRule;
      eho:hasFormat :sparqlQueryFormat ;
      eho:hasRuleURL <http://hypios.com/expert-search/example-rules/rule_1_1> ;
      eho:hasParameter :p1_1.
:sparqlQueryFormat  a  eho:ExchangeFormat ;
      eho:specificationDocument <http://www.w3.org/TR/rdf-sparql-query/> .
:p1_1 a eho:RuleParameter ;
      eho:parameterType rdf:Resource ;
      eho:parameterName "expertise topic" .
```

Code 1. An example of an expertise hypothesis metadata expressed in EHO

3   Recommendation of Expertise Hypotheses Based on Linked Data Metrics

We propose a set of Linked Data metrics aimed at guiding the expert finding process and facilitating the selection of appropriate expertise hypotheses for a given topic of expertise. The basic assumption behind our metric-based approach is that the presence and frequency of certain types of user traces as well as the presence and frequency of domain specific topics in the linked data of a particular expertise community may indicate some patterns of user behavior in that community. In other words, the constitution of linked data could reveal that certain types of traces are more significant for expertise detection related to certain topics of expertise. This assumption is based on the observed difference in the communication media that are dominant in different expertise communities. This difference becomes obvious if we look at the most recent studies of scientific communications in different fields. While biology-focused studies struggle to explain dissemination on blogs and online

---

[3] http://ontologies.hypios.com/eho

encyclopedia [10] and do not even try to look at Twitter, studies focused on Semantic Web research community can produce understandings of adoption of Twitter as a prominent tool in the scientific communication [11]. Being based on the above stated assumption, our Linked Data metrics enable the detection of dominant user trace types for a particular topic of expertise. Specifically, their role is to quantitatively estimate the extent of relation between a particular type of user trace and a particular topic of expertise, based on the observed constitution of the given linked dataset. Once the dominant trace type is identified for a given topic of expertise, we select expertise hypotheses that use this dominant type of trace and propose them to the user who wants to perform an expert search. The user can then select the particular expertise hypothesis of his preference, and run the expert search on a linked dataset of his choice (see Section 4).

In the following sub-sections we first motivate the introduction and use of our Linked Data metrics (Section 3.1); subsequently, we present these metrics (Section 3.2), then the experiment, which we conducted in order to evaluate the metrics (Section 3.3), and finish the section with the discussion on the scope of the metrics (Section 3.4).

3.1 Motivation

The main motivation behind our Linked Data metrics-based approach is the following:

- Take advantage of the growing number of Linked Data sources, containing diverse and constantly emerging kinds of user traces, to construct a flexible and versatile expert finding approach. Linked Data could prove especially convenient for integrating experts' contact data from different sources, as well as for taking into account the similarities between topics of expertise;
- Base the recommendation of expertise hypotheses for a given domain on the metadata (descriptions and statistics) of linked dataset(s). As opposed to the possibility of deriving overly general and potentially outdated recommendations from the analysis of the expert search literature and from qualitative studies, the metric-based approach enables recommendation that would change together with the changes of human practices and data patterns;
- In cases where the dataset for expert finding is not yet chosen, we want to provide a way to use global LOD statistics to detect global data patterns, and thus suggest the appropriate user trace types to use in the expert search approach. Based on the suggested user trace types, searching for linked open datasets that contain relevant data should be feasible.

3.2 Designing Linked Data Metrics

We have designed a number of metrics to quantify the relation between a certain user trace type and a certain topic of expertise [12]. Although a number of those metrics gave promising results in our experiments, their calculation proved to be overly time-consuming and the calculation process over multiple distributed datasets proved as almost impossible. We thus decided to take into account only the metrics that can be calculated from the dataset descriptions, without having to run additional SPARQL queries over the data. Here we mostly refer to the growing practice of providing

dataset descriptions using the VoID[4] ontology and the dataset statistics using the SCOVO[5] ontology. In addition to this practice, the approaches to add data summaries to data storage systems in order to increase the performance of SPARQL queries (e.g., [13][14]) might also provide the data about the dataset composition that could be used to calculate the metrics.

3.2.1 Metrics Based on Data Quantity

The simplest imaginable metrics are based on the quantity of available data of a certain user trace type. We define Q_t to be the number of available instances of type t. A natural question to ask is what this number is relative to: a dataset, or the total LOD? We call this scope of metric calculation, and it can be different in different cases of metric use. We detail the question of metric calculation scope in Section 3.4. Further on, it would be interesting to know the number of instances of a certain type satisfying some condition (e.g., having some particular concept as value of the dct:subject property). We thus define $Q_{t,\,C}$ where C^6 is a set of concepts (topics) that are associated with the instances to be counted. Although they may be considered as metrics themselves, Q_t and $Q_{t,C}$ serve mostly as a ground for defining more complex metrics based on topic distribution (Section 3.2.2).

3.2.2 Metrics Based on Topic Distribution

We assume that a considerable co-occurrence of particular topic with particular type of trace instances could successfully drive search for experts on the given topic. The <trace type, topic> patterns that would be found in this way could be used to privilege expertise hypotheses that rely on the particular trace type.

We define *subject homogeneity* $SH_{t,s}$ as number of user trace instances of type t that are associated with topic s, divided by the total number of user trace instances of type t. Subject homogeneity shows the degree of presence of the subject s within user traces of the type t. We also define *type homogeneity* $TH_{t,s}$ as number of user trace instances of type t that are associated with topic s, divided by the total number of user trace instances associated with topic s. This metric shows the ratio of use of particular trace type with instances relevant for a particular topic. At the same time $TH_{t,\,s}$ represents the upper bound of recall for expertise hypothesis using trace type t and searching for expert on topic s.

$$SH_{t,s} = \frac{Q_{t,s}}{Q_t} \quad TH_{t,s} = \frac{Q_{t,s}}{Q_{out:UserTrace,s}} \tag{1}$$

3.3 Measuring the Correlation Between Values of Metrics and the Performance of Expert Search

In this section we present the experiment we conducted in order to determine if there is a correlation between the values produced by our metrics and the performance of the expert search.

---

[4] http://vocab.deri.ie/void/
[5] http://sw.joanneum.at/scovo/schema.html
[6] When C is a set composed of only one topic concept s, then it is simply called s.

3.3.1   The Experimental Setting

In order to perform the experiment, we needed a dataset that would contain different types of user traces (e.g., blog and microblog posts, publications). Since we were not able to find a single LOD dataset that meets these requirements, we needed to combine data from several LOD datasets and assemble a sample dataset for testing purposes. After creating the sample dataset, we conducted a user study in order to identify in this dataset experts on a chosen set of topics. Specifically, we chose three expertise topics: dbpedia:Linked_Data, dbpedia:SPARQL, and dbpedia:Open_Data. The study ended by generating a "gold standard" for experts in the considered expertise topics. The gold standard actually consisted of a group of recognized experts for each considered topic. Having this data, we have calculated the correlation between the values produced by our Linked Data metrics (calculated based on the sample dataset constitution) and the performance of expert search that could be achieved by applying particular kinds of expertise hypotheses (on the same sample dataset).

3.3.1.1 Creation of the Sample Dataset In order to build the sample dataset we took dumps from the Semantic Web Dog Food[7] dataset, containing mostly conference and workshop publications. Additionally, we queried Sindice.com for all instances of the type sioct:BlogPost. This produced a total of 1436 instances of swrc:Publication and 837 instances of sioct:BlogPost. We used our tool for exposing semantics of tweets as linked data [15] to produce a dataset of archived tweets from the latest conferences related to the domain of Semantic Web and Web of Data. The conference archives came from the Twitter archiving service TwapperKeeper.com, and resulted in 6631 tweets in RDF. We also used the SlideShare2RDF[8] service to obtain 1657 slideshows relevant to the Web-related keywords. Such constitution of our sample dataset makes it representative for the domain of Web of Data and Semantic Web, since it contains almost all user traces (publicly available on the Web) of the mentioned trace types that were findable with a reasonable effort, on the dataset creation date which is 2010-06-15. Being domain focused, our experiment produced results that are directly applicable to the selected domains of expertise (Web of Data and Semantic Web). To make the results more broadly applicable, we intend to repeat the experiment with datasets representative for some other domains.

3.3.1.2 Data Cleaning. We faced many issues of data quality in the data retrieved from LOD, so we had to correct some data and add missing elements. In particular, the instances of sioct:BlogPost tend to be irregular. Many of them were missing the author information, so we had to construct a foaf:Agent instance based on the URL of the blog. Slideshare data was of type sioc:Item, which is quite broadly defined, so we added the bibo:Slideshow as additional type. Next problem was that only tweets produced by our system had DBpedia concepts associated with them as topics. Publications sometimes had them, but mostly had only textual topics; the same was with slideshows. None of the blog posts had them. Since our approach heavily relies on the availability of topics, we enriched the instances with topics where they were

---

[7] http://data.semanticweb.org/
[8] http://linkeddata.few.vu.nl/slideshare/

missing. This was done by running Zemanta[9] concept extraction service over the textual data related to instances (mostly values of dc:description, dc:title, sioc:content and swrc:abstract). The concepts that were obtained in this way were added to the instances using dct:subject property.

3.3.2 Expert Search Performance Measures

The performance of expert search is most commonly measured by precision and recall. By *precision* we understand the ratio of true positives, i.e. true experts in the total number of found expert candidates. By *recall* we understand the number of true experts found divided by the total number of true experts in a given domain. In our case determining the total number of true experts in a given domain is on the boundary of impossible, and thus we define an adapted measure of recall relative to the possibilities of a dataset (or the set of datasets) used for expert search. *Relative recall* of a particular dataset is the number of true experts found divided by the total number of true experts findable in that dataset. We consider that an expert is not findable in a dataset if the dataset's graph does not contain any path from the expert node to the node representing the topic of interest in the expert search process. Finally, since we want to favor the expert search approaches that lead to a good balance in precision and relative recall, we use the *F'*-measure that represents a harmonic mean between precision and relative recall. Our version of the F-measure gives equal importance to precision and relative recall.

$$F' = 2 \frac{precision \bullet relative\_recall}{precision + relative\_recall} \qquad (2)$$

3.3.3 The Gold Standard Creation

In order to evaluate the metrics in terms of their correlation with the expected performance of the expert search, we first needed to identify the true experts that may be found in our sample dataset i.e., we needed to create the gold standard. To obtain this information we have conducted a user study in which four expert users evaluated the potential experts from the sample dataset. The users were all master students at prestigious universities, who have a record of doing some important work (internship, project, bachelor thesis, etc.) on the topic of interest. The evaluation was done using an iterative process inspired by the Delphi method [16] to achieve agreement of experts about a certain opinion (most commonly a prediction). Experts give their predictions and justify them. In the second round they read the predictions and justifications of other experts, and can change their mind and thus agree on a common prediction. We used the same process to reach an agreement about the true experts on a given topic. To calculate the agreement levels between evaluators we used the kappa (*k*) agreement metric, defined in [17], to calculate the rater agreement in each phase. The *k*-statistic measures the chance corrected agreement between raters, using the confusion matrix shown in Table 1. Using these set definitions the *k*-statistic is calculated using the following formula:

$$\kappa = \frac{2(ad - bc)}{(a+c)(c+d) + (b+d)(a+b)} \qquad (3)$$

---

[9] http://zemanta.com

Table 1. Rater Agreement Confusion Matrix

		Rater 1	
		Positive	Negative
Rater 2	Positive	a	b
	Negative	c	d

We have first extracted the lists of findable expert candidates for each of the three topics (dbpedia:Linked_Data, dbpedia:SPARQL, and dbpedia:Open_Data.), i.e., all expert candidates in the sample dataset for which there was any path between the expert candidate node and the topic node in the data graph. The evaluators were given these lists of candidate experts. A dereferenceable URI was provided for each expert candidate, so that the evaluators could retrieve additional information about the expert candidate and judge his expertise. Through the URI they could gain access to the expert candidate's publications, tweets, slides, blog posts, or personal homepage. Based on their judgment they assigned mark "0" or "1" to an expert candidate based on their perception of the candidate's real expertise ("1" for true experts, and "0" otherwise). The values of rater agreement statistic after the first round were very low (see the second column of Table 2). We thus conducted the second round, in which the raters could reach an agreement by correcting their results and commenting upon the other raters' marks. After the second round, the rater agreement for all three topics was above the 0.6 threshold which allowed us to take the rating from the second round as our gold standard.

Table 2. Average Rater Agreement Between Each Two Rater Pairs in the First and Second Round

	First Round	Second Round
Linked Data	0.4215	*0.6878*
SPARQL	0.4673	*0.6482*
Open Data	0.4673	*0.7228*

3.3.4 Measuring Correlation

For each type of user trace from our sample dataset, we have calculated the precision and relative recall that could be obtained if a hypothesis that uses this type of trace was applied on the sample dataset. The calculation is based on the assumption that the least restrictive hypothesis is used for a particular type (like the one accepting only one trace type to be enough for considering its author as expert).

After conducting the Pearson correlation test on the given data, we have obtained the positive values for correlation between TH and relative recall (r=0.846), between TH and F' (r=0.778), and between SH and precision (correlation coefficient r=0.619). Since all our values are above the significance threshold (r=0.576 for our sample size), we can consider the results to be statistically significant. As for the basic measures, Q_t shows no correlation with expert search performance measures, and $Q_{t,C}$ behaves like TH, just with slightly weaker correlation (r=0.777 with relative recall and r=0.761 with F'). This conclusion allows us to ground the expertise hypothesis recommendation on the values of SH and TH. In other words, based on those two metrics we would suggest the trace type that is expected to produce the best results (precision or relative recall), and the user could then choose among the hypotheses

Table 3. Values of Metrics and Expert Search performance

t	Q_t	Q_{tc}	SH	TH	precision	relative recall	F'
Linked Data							
sioc:BlogPost	837	10	0.083	0.024	0.800	0.050	0.094
sioc:MicroblogPost	6631	96	0.014	0.236	0.469	0.296	0.363
bibo:Slideshow	1657	55	0.033	0.135	0.673	0.243	0.357
swrc:Publication	1436	86	0.060	0.211	0.721	0.408	0.521
SPARQL							
sioc:BlogPost	837	27	0.030	0.218	0.259	0.206	0.229
sioc:MicroblogPost	6631	13	0.002	0.105	0.385	0.147	0.213
bibo:Slideshow	1657	33	0.020	0.266	0.303	0.294	0.298
swrc:Publication	1436	29	0.020	0.234	0.414	0.353	0.381
Open Data							
sioc:BlogPost	837	5	0.006	0.013	0.600	0.025	0.048
sioc:MicroblogPost	6631	45	0.007	0.116	0.511	0.192	0.279
bibo:Slideshow	1657	80	0.048	0.207	0.300	0.200	0.240
swrc:Publication	1436	150	0.105	0.399	0.666	0.583	0.518

that rely on this trace type. Apart from expertise hypotheses suggestion, this conclusion can serve as a ground for semi-automatic hypothesis generation in which most promising user trace types would be suggested to the user who may combine them to obtain an optimal hypothesis.

3.4 The Scope of Metric Calculation

Although in our experimental setting we calculate the values of Linked Data metrics on only one dataset, it is fully imaginable to calculate them on different levels. Possible scopes of metric calculation are: (1) one dataset, (2) set of datasets and (3) the global LOD. All scopes have their own advantages, and disadvantages – many of which can be remedied.

3.4.1 Calculating the Metrics on One Dataset
Calculation of metrics for a dataset we control might be the easiest and the most accurate option, since we can normally control the generation of indexes or statistical data related to the dataset, and make sure they are accurate. A potential drawback of using this scope is that the chosen dataset might not be representative of the overall data (the portion of particular trace types in the data set might significantly differ from the global one). A possible remedy in this case is to estimate the overall structure of the LOD and then (a) make sure the dataset is representative, or (b) compose several datasets in order to reach representativeness.

3.4.2 Calculating the Metrics on a Set of Datasets
In order to calculate the metric values on the scope of several datasets, we could integrate their VoID and SCOVO descriptions to gain the knowledge about the overall composition of our chosen scope. However, at present, relying on indexes and descriptions to apply the Linked Data metrics has several potential limitations due to the data quality issues in LOD – something that has been pointed out in many sources, most notably [18] [19]. The most relevant data quality issues for our metrics are the following: 1) missing explicit type declarations of LOD instances, and 2) missing

topic information of user traces contained in LOD. As shown in the creation of our sample dataset, those issues can be remedied on the dataset level, but not on the level of pre-calculated dataset descriptions that would simply reflect incomplete data. Thus relying solely on dataset descriptions to calculate the metrics on the level of the global LOD might involve some risks on accuracy, and the metric values obtained on this scope would be just approximations of real values. Dataset ranking [20] and quality filters [21] could help select the dataset descriptions to take into account and thus choose a scope that would grant higher accuracy.

3.4.3   Calculating the Metrics on the Global LOD

In cases where the dataset to work with is not chosen in advance, the global LOD might be the only choice for the scope of metrics. The same estimation strategy as for a set of datasets can be used, by relying on VOID + SCOVO descriptions. Systems[10] that automatically calculate and expose dataset descriptions could help obtain the statistics on the LOD level. The second way to perform the LOD-level estimation is to query a Semantic Web indexing service, such as Sindice.com, and obtain the number of known instances of a certain trace type, as well as traces related to a certain topic, that would allow us to calculate the metrics. This method suffers from the same dependence on the accuracy and completeness of the data in the index as reported above (Section 3.2.2).

4   hy.SemEx – The Expertise Hypothesis Recommendation and Expert Search System

Based on our findings about recommending expertise hypotheses, we have created a proof of concept system that supports a user in finding the right expertise hypothesis, and then using it to find experts. The system, currently in private alpha testing, is intended to work as a support to the Open Innovation process at hypios.com where it would allow innovation seekers to identify experts on topics of their interest and send them open innovation problems. In the following sub-sections we describe the system and present the feedback obtained from our alpha testers.
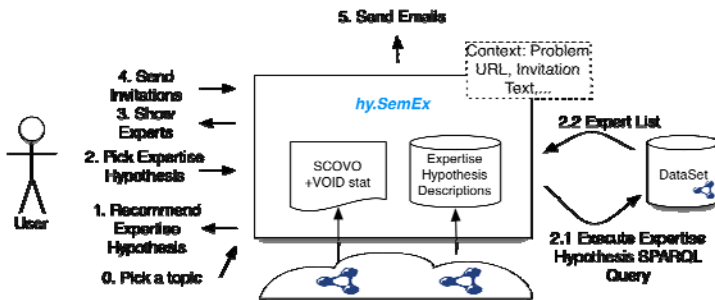

Fig. 1. Scenario of use of hy.SemEx

---

[10] https://www.hpi.uni-potsdam.de/naumann/sites/btc2010/

4.1   The Context and Scenario of Use

For the purpose of providing an expert search service to its clients, hypios harvests structured data containing user traces (mostly publications), and saves them in a 4store[11] triple store. In addition, a tool called SolverSurfer crawls the Web and processes free text sources containing user traces. The extracted data is then fed to the common triple store. This process makes the store resourceful for at least the domains relevant to the open innovation problem of client's interest. Hy.SemEx recommends the expertise hypotheses that client should use to find experts in this base, and allows her to run the expert search (execute the expertise hypothesis). The client can later send messages about his open innovation challenge to the identified experts. A custom designed library called SilverRabbit is in charge of sending messages over various channels of communication (tweets, e-mails, comments on slideshare presentation, etc.) depending on the known user accounts and contact data of the candidate experts. In our basic scenario of use (Figure 1), the user (innovation seeker) is offered with the list of topics, potentially relevant to her open innovation problem. The user then chooses one topic to perform the expert search and enters its DBpedia URI in the hy.SemEx system (Step 0 on Figure 1). She then expresses her preference towards a performance measure (precision or recall) and invokes the expertise hypothesis recommendation. Based on the user's choice, the appropriate metric is calculated, using the statistics about the hypios triple store as a scope of metric calculation. Using the metric values, the most appropriate user trace type is selected. The system then finds all the expertise hypotheses that use this trace type, and proposes them to the user (1) who picks one of them (2). The SPARQL query corresponding to the chosen hypothesis is then run (2.1) and the list of found experts is presented to the user (3). If the user is happy with the list, she can send it to the SilverRabbit messaging component (4), which sends the invitation (5) to the experts to come and solve the problem. The messages have previously been customized for the problem in question as a part of the open innovation campaign set-up paid by the user. Additional information about the system, such as screen-shoots, expertise hypotheses used, etc. is provided on our website[12].

4.2   Implementation Details

The system is designed for flexibility. In the core of the system is a Jena triple store that stores RDF descriptions of expertise hypotheses known to the system. New expertise hypotheses described using the EHO ontology can be easily fed to the system. The system currently works with the hypios triple store as a source for user traces and calculates the metrics based on this store as a scope. However, it is possible to reconfigure the system (in a configuration file) to calculate the metrics based on global statistics from Sindice.com. Similarly, the system can be plugged on any other data store containing traces, where it would work in a similar way it works on the hypios triple store. The endpoint information as well as dataset statistics can be fed to the system in an RDF file containing VOID and SCOVO-based descriptions and

---

[11] http://4store.org/
[12] http://research.hypios.com/?page_id=142

statistics[13]. In order to support statistics related to the number of traces of a certain type, as well as the number of traces with a certain topic, we have extended SCOVO with two additional properties[14]. The system is fully built in Java, using Jena to work with RDF data, and using Tapestry as a Web application framework.

4.3   Evaluation

Since the success of expert search depends on the choice and quality of dataset as much as on the choice of appropriate hypothesis, it is difficult to evaluate the quality of expertise hypothesis suggestions based on the final lists of found experts and users' satisfaction with them. For this reason we decided to directly evaluate users' satisfaction with the hypotheses suggested by the hy.SemEx system. To do that, we have conducted an evaluation study with 10 participants, and follow-up interviews with 3 of them who accepted to do it. The study participants have tested the system for 7-10 topics of their own expertise, and rated the suggested class of expertise hypothesis[15] on a scale of 1-5 (1: this class of expertise hypotheses should not be used for finding experts on this topic; 5: this is the best class for this topic). For each topic, the participants were asked to evaluate how well the proposed class of expertise hypotheses allows for finding experts on the given topic if the focus of expert search was finding: (a) the best experts, (b) as many experts as possible on the given topic. These two variations focused on: (a) precision as the desired criterion for hypothesis recommendation (where hy.SemEx used SH metric), and (b) relative recall as the recommendation criterion (corresponding to TH metric). This gave a total of 176 evaluations. The fact that the study participants chose the topics for which they have substantial knowledge, assured that they are aware of the type of communication media which is predominantly used in the community to share knowledge; this further makes their ratings reflective of the correctness of suggestions given by hy.SemEx.

The average participants' ranking for the case (a) where precision was favored was 4.234 ± 0.857 and 3.947 ± 0.751 in case (b) where recall was preferred. This result shows that suggesting expertise hypotheses to the users based on the SH and TH metrics gives reasonably good results. Specifically, the class of expertise hypothesis recommended to the participants by hy.SemEx was based on the type of user trace that was first ranked by our Linked Data metrics (SH in case (a) and TH in case (b)). We thus have shown, not only the overall correlation demonstrated in Section 3.3.4, but a considerable fitness of best-ranked type of user trace (and the corresponding class of expertise hypotheses).

After receiving suggestions, the participants could pick one of the suggested expertise hypotheses, or alternatively say they would prefer another hypotheses from a list of all hypotheses we know about[13]. The participants preferred different hypotheses for precision and different for recall. For instance, within the class of hypotheses that relied on blogs, the most preferred were the expertise hypothesis taking only authors of blog posts with more than 5 comments; however, when recall

---

[13] An example data set meta data file, along with a complete list of hypotheses used is provided on `http://research.hypios.com/?page_id=142`

[14] `http://ontologies.hypios.com/traceStat`

[15] We consider that the hypotheses using the same type of user trace belong the same class of hypotheses (identified by the type of user trace in question).

was demanded, they preferred the hypothesis that considered all the authors of at least two blog posts as experts. Similar patterns were observed with all classes of expertise hypotheses. In the follow-up interviews, the participants explained this by the need to use more restrictive hypotheses when precision is required. This indicates the need to include the information about restrictiveness in expertise hypothesis metadata (i.e., in the EHO ontology) and allow for automatic prioritization of hypotheses according to the user's preference towards precision or recall. Along with usability improvements, the participants also expressed a requirement for being able to change the numeric parameters of hypotheses (e.g., the number of traces taken as threshold); as well as for knowing which hypotheses were most used by others before.

5   Related Work

Although to our knowledge there are no similar approaches to use Linked Data metrics for expertise hypothesis recommendation, or for improving expert search in general, there are some similar metric-based approaches worth mentioning. The approach of Tran et al. [22] that uses data summaries to construct query routing plans for distributed query execution on LOD is conceptually very similar to ours, but applied for a different purpose. We are also not the first to use LOD as source for expert finding, but the existing approaches mostly remain mono-hypothesis i.e., limited to particular types of user traces, and even to particular topics of expertise. Saffron [23] for instance uses the Semantic Web Dog Food[16] data about publications, and augments it with research topics mined from the text, in order to allow for browsing of experts – the system does not support multiple user trace types.

6   Conclusions and Future Work

In this paper we discussed novel opportunities for expert search opened by the appearance of user traces in the form of Linked Data, most notably the opportunity of letting users choose appropriate expertise hypotheses for the relevant topic of expertise. We examined how Linked Data metrics, that reveal the constitution of a linked dataset (or set of datasets), could help to detect a good type of user trace to use for expert finding, and thus help the user prioritize those expertise hypotheses that rely on this particular type of trace. Our gold standard-based study showed that there is a significant correlation between our metrics and the expected precision and relative recall of expert search based on particular user trace types. We presented an application of this finding in a proof of concept system hy.SemEx, which uses Linked Data metrics to recommend expertise hypotheses to the user. We have evaluated the user satisfaction with the recommendations, and have found it to be very positive for the first-ranked hypothesis. In addition, the recommendation of hypothesis is not the only possible use of metrics. Approaches for helping the user construct new expertise hypothesis could also be backed up with Linked Data metrics.

In the future work we intend to repeat our experiment on a dataset that would reflect another domain and see if the same metrics still work well. Another future

---

[16] http://data.semanticweb.org

work direction is to look at other characteristics of expertise hypotheses (e.g., their restrictiveness), and not just the user trace type, and verify if there is a way to base the hypothesis recommendation on them as well. Finally, we will observe the functioning of hy.SemEx and collect the data needed to evaluate the actual expert search and not just the hypothesis recommendation part. We also intend to make the system user-friendlier, especially by making it more configurable, and by hiding the technical aspects such as concept URIs, from the user.

References

1. Chesbrough, H.W.: Open Innovation: The New Imperative for Creating and Profiting from Technology. Harvard Business Press, Boston (2003)
2. Buitelaar, P., Eigner, T.: Topic Extraction from Scientific Literature for Competency Management. In: The 7th International Semantic Web Conference, Karlsruhe (2008)
3. Kolari, P., Finin, T., Lyons, K., Yesha, Y.: Expert Search using Internal Corporate Blogs. In: Workshop on Future Challenges in Expertise Retrieval, SIGIR 2008, pp. 2–5 (2008)
4. Chua, S.J.: Using web 2.0 to locate expertise. IBM Centre for Advanced Studies Conference (2007), Retrieved from
   `http://portal.acm.org/citation.cfm?id=1321250`
5. Demartini, G.: Finding experts using wikipedia. In: Proceedings of the Workshop on Finding Experts on the Web with Semantics; ISWC/ASWC 2007, Busan, South Korea (2007)
6. Noll, M.G., Yeung, C.A., Gibbins, N., Meinel, C., Shadbolt, N.: Telling Experts from Spammers: Expertise Ranking in Folksonomies. In: Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, Boston, MA (2009)
7. Adamic, L., Zhang, J., Bakshy, E., Ackerman, M.: Knowledge sharing and yahoo answers: everyone knows something. In: Proceedings of the 17th International Conference on World Wide Web, Beijing, China, pp. 665–674. ACM, New York (2008)
8. Becerra-Fernandez, I.: Searching for experts on the web: a review of contemporary expertise locator systems. ACM Trans. on Internet Technology 6(4), 333–355 (2006)
9. Stankovic, M., Wagner, C., Jovanovic, J., Laublet, P.: Looking For Experts? What can Linked Data do for You? In: Pre-proceedings of Linked Data on the Web 2010 (LDOW) Workshop, within WWW 2010 Conference, Raleigh, NC, USA, April 26-30 (2010)
10. Comité Consultatif National d'Éthique pour les Sciences de la Vie et de la Santé (France), Communication d'informations scientifiques et médicales et société: enjeux éthiques.-Mars (2010), Retrieved from
    `http://www.upf.edu/pcstacademy/_docs/CCNE-Avis109.pdf`
11. Letierce, J., Passant, A., Breslin, J., Decker, S.: Understanding how Twitter is used to spread scientific messages. In: Proc. of the Web Science Conference, Raleigh, NC, USA, April 26-27 (2010)
12. Stankovic, M.: Open Innovation and Semantic Web: Problem Solver Search on Linked Data. In: Proc. of International Semantic Web Conference 2010, Shanghai, China, November 7-11 (2010)

13. Khatchadourian, S., Consens, M.: Exploring RDF Usage and Interlinking in the Linked Open Data Cloud using ExpLOD. In: Proc. Linked Data on the Web Workshop 2010 on the WWW 2010, Raleigh, NC, vol. 1430 (2010)
14. Harth, A., Hose, K., Karnstedt, M., Polleres, A., Sattler, K.: Data Summaries for On-Demand Queries over Linked Data. In: Proceedings of the 17th International Conference on World Wide Web, WWW 2010, pp. 411–420. ACM Press, Raleigh (2010)
15. Stankovic, M., Rowe, M., Laublet, P.: Mapping Tweets to Conference Talks: A Goldmine for Semantics. In: Proceeding of the Third Social Data on the Web Workshop SDoW 2010, Collocated with the International Semantic Web Conference, Shanghai, China, November 8 (2010)
16. Rowe, Wright: The Delphi technique as a forecasting tool: issues and analysis. International Journal of Forecasting 15(4) (October 1999)
17. Fleiss, J.L.: Statistical methods for rates and proportions, 2nd edn. John Wiley, NY (1981)
18. Hogan, A., Harth, A., Passant, A., Decker, S., Polleres, A.: Weaving the Pedantic Web. In: 3rd International Workshop on Linked Data on the Web (LDOW 2010) at WWW2010. CEUR Workshop Proceedings, vol. 628, CEUR-ws.org (2010)
19. Jain, P., Hitzler, P., Yeh, P.Z., Verma, K., Sheth, A.P.: Linked Data Is Merely More Data. In: Brickley, D., Chaudhri, V.K., Halpin, H., McGuinness, D. (eds.) Linked Data Meets Artificial Intelligence. Technical Report SS-10-07, pp. 82–86. AAAI Press, Menlo Park (2010)
20. Toupikov, N., Umbrich, J., Delbru, R., Hausenblas, M., Tummarello, G.: DING! Dataset Ranking using Formal Descriptions. In: WWW 2009 Workshop: Linked Data on the Web (LDOW, Madrid, Spain (2009), Retrieved from `http://sw-app.org/pub/ldow09-ding.pdf`
21. Bizer, C., Cyganiak, R.: Quality-driven information filtering using the wiqa policy framework. Journal of Web Semantics: Science, Services and Agents on the World Wide Web 7(1), 1–10 (2009)
22. Tran, T., Zhang, L., Studer, R.: Summary Models for Routing Keywords to Linked Data Sources. In: Proceedings of the 9th International Semantic Web Conferene 2010, pp. 1–16. Springer, Shanghai (2010)
23. Bordea, G.: Concept Extraction Applied to the Task of Expert Finding. In: Extended Semantic Web Conference 2010 PhD Symposium, Heraklion, Grece (2010)