CRAWL·E: Distributed Skill Endorsements in Expert Finding

Sebastian Heil, Stefan Wild, and Martin Gaedke

Technische Universität Chemnitz, Germany
firstname.lastname@informatik.tu-chemnitz.de

Abstract. Finding suitable workers for specific functions largely relies on human assessment. In web-scale environments this assessment exceeds human capability. Thus we introduced the CRAWL approach for Adaptive Case Management (ACM) in previous work. For finding experts in distributed social networks, CRAWL leverages various Web technologies. It supports knowledge workers in handling collaborative, emergent and unpredictable types of work. To recommend eligible workers, CRAWL utilizes Linked Open Data, enriched WebID-based user profiles and information gathered from ACM case descriptions. By matching case requirements against profiles, it retrieves a ranked list of contributors. Yet it only takes statements people made about themselves into account. We propose the CRAWL·E approach to exploit the knowledge of people *about* people available within social networks. We demonstrate the recommendation process for by prototypical implementation using a WebID-based distributed social network.

Keywords: Endorsements, Expert Finding, Linked Data, ACM, WebID.

1 Introduction

Knowledge work constitutes an ever increasing share of today's work. The nature of this type of work is collaborative, emergent, unpredictable and goal-oriented. It relies on knowledge and experience [17]. Traditional process-oriented Business Process Management (BPM) is not well applicable to areas with a high degree of knowledge work [21]. Addressing this issue, non-workflow approaches [5], in particular Adaptive Case Management (ACM), gain more relevance [8].

ACM systems assist knowledge workers. They provide the infrastructure to handle dynamic processes in a goal-oriented way. Traditional BPM solutions feature a-priori processes modeling. Contrary to them, ACM systems enable adaptivity to unpredictable conditions. Uniting modeling and execution phases contributes accomplishing this adaptivity. A case represents an instance of an unpredictable process and aggregates all relevant data. For adapting it to emergent processes, case owners can add ad-hoc goals. There are cases where persons currently involved cannot achieve all goals. In these cases it is necessary to identify suitable experts based on the skills and experience required for that particular part of work.

S. Casteleyn, G. Rossi, and M. Winckler (Eds.): ICWE 2014, LNCS 8541, pp. 57–75, 2014.

The nature of knowledge work often implies cross-enterprise collaboration. It necessitates access to information about the persons involved, e.g., CV and contact data. It is unlikely that all potential collaborators use the same social network platform for storing personal information. Cross-platform relationships are hard to follow. Such "walled gardens" [1] would complicate expert finding. Distributed social networks are well-suited for the knowledge work domain. Companies or knowledge workers can host their own profiles. The profile can include work experience and skill information. Interlinking these distributed profiles establishes a social network. Such social network could overcome the data silo characteristic of walled gardens. This would enable crawling the network to identify experts.

Finding suitable workers for specific functions largely relies on human assessment. Assessors have to make their decisions depending on the requirements at hand. This decision making requires knowledge of potential contributors and their experience. The selection complexity increases with the amount of eligible contributors and work requirements. Human assignment does not scale well, especially not with web-scale processes [12]. Often work is assigned to workers who are not the most suitable experts available. This can cause mediocre outcomes and longer times to completion. Dealing with this problem requires software support for finding and addressing knowledge workers to contribute to cases.

In [15] we introduced CRAWL, an approach for Collaborative Adaptive Case Management. It leverages various Web technologies to automatically identify experts for contributing to an ACM case. CRAWL recommends a set of eligible workers. It uses Linked Open Data, enriched WebID-based user profiles and information gathered from project or case descriptions. We created a vocabulary to express the skills available and the skills required. It extends user profiles in WebID-based distributed social networks and case descriptions. CRAWL's semantic recommendation method retrieves a ranked list of suitable contributors whose worker profiles match the case requirements.

Problem. The skill information about a person is limited to the expressive power and will of this particular person. As a consequence, CRAWL only takes statements people made about themselves into account. Such statements, however, might be unspecific, exaggerated or even wrong. This affects the expert finding process and makes the assessor's task more difficult and time-consuming.

There are three possible kinds of statements about skills, as shown in Figure 1. The most basic form are *skill self-claims*, statements by someone claiming that he himself has a certain skill. *Skill assignments*, on the other hand, are statements by someone claiming that someone else has a certain skill. Statement claimed by someone for himself and confirmed by someone else are *skill affirmations*. We refer to these three kinds of skill statements together as *skill endorsements*.

With knowledge work increasingly becoming an important and widespread part of work [9] and ACM evolving as an approach addressing this type of work, we are convinced that enabling knowledge workers to find the right collaborators to contribute to multi-disciplinary cases impacts the performance of future enterprises [5]. The value add by skill endorsements will trigger a demand to incorporate them into distributed worker profiles and expert finding algorithms.

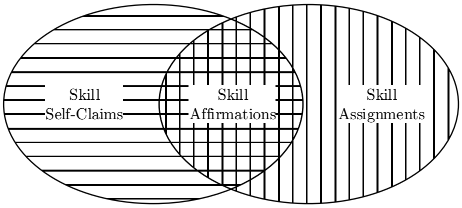

Fig. 1. Three kinds of skill endorsements

Overall Objective. To exploit the knowledge of people *about* people available within social networks, we aim at integrating skill assignments and skill affirmations in addition to skill self-claims into the distributed expert finding process.

Contributions. To contribute to the overall objective, we must achieve the following objectives to fully use skill endorsements in distributed social networks:

1. To enable expert finding in distributed social networks
2. To increase credibility of skill self-claims
3. To allow assigning skills the endorsee did not consider
4. To prevent unwanted skill endorsements
5. To express skill endorsements in distributed user profiles
6. To incorporate skill endorsements in distributed expert finding
7. To facilitate a differentiated consideration of skill endorsements

The paper is organized as follows: Section 2 illustrates the necessary objectives in order to achieve the overall objective. The background is provided in Section 3. We present the CRAWL·E approach in Section 4. Section 5 evaluates the approach. We discuss work related to ours in Section 6 and conclude the paper in Section 7.

2 Objectives of Distributed Expert Finding

This section describes the objectives in greater details. To illustrate the need for achieving each objective, we use different personae. All of them are knowledge-workers and members of a distributed social network. They have a different character and pursue different goals. Figure 2 shows the corresponding social graph. Black solid arrows indicate knows-relationships, blue dotted arrows symbolize endorsements. The personae are characterized in the following:

Alice wants to record her skills. She likes to include all skills from her current job, past jobs and education. Alice intends to record them in a way others can easily access them. She does not want to spend too much effort in achieving this goal.

Bob is a co-worker of Alice. He knows Alice very well because he worked together with her in many projects. Bob trusts Alice and Alice trusts him.

Casey is a case owner who wants to find and recruit the best persons for a job.

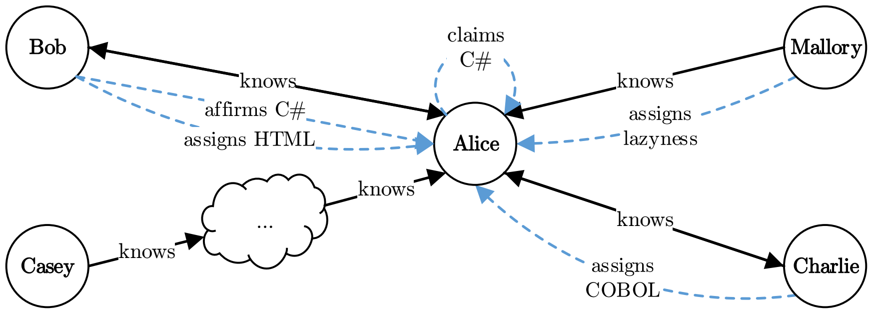

Fig. 2. Social Graph with Endorsements

Charlie is another co-worker of Alice. Compared to Bob, he is not that close to Alice. Charlie worked together with Alice in only one project long time ago.

Mallory is a bad guy. He dislikes Alice and wants to damage Alice's reputation.

Having described the personae that are used throughout this paper, we continue with outlining the objectives.

Objective 1: To enable expert finding in distributed social networks. Interoperability, compatibility, and portability of skill information is no issue in conventional centralized social networks like LinkedIn. Expert finding benefits from the (virtually) monolithic data layer of such social networks. By contrast, this does not exist in distributed social networks which are formed by interlinked, distributed profiles[1]. Therefore, skill endorsements in profiles are also inherently distributed over the network. This requires to ensure discovery, comparability and description of skills across organization boundaries. Each skill endorsement needs to be described properly to facilitate comprehension and avoid misunderstandings. To allow adding further skill information, both skill set and skill descriptions need to be accessible, extensible and linkable. So, persons can easily refer to descriptions of the skill endorsements they made. When Alice claims she has a skill, this endorsement must be associated to the person making the claim i.e., to her. This is necessary because information a person produces, belongs to her. To persist skill endorsements associated to the person stating them, they need to be stored and connected with the person's identity. This enables persons and machines to detect skill endorsements, provided that relevant data is accessible in an easy-to-process manner. To achieve this objective, we need to deliver 1) an extensible description for each skill, 2) a way to attach all kinds of skill endorsements to persons, 3) a place to store each person's skill endorsements, and 4) a procedure to find experts in distributed social networks based on skills.

Objective 2: To increase credibility of skill self-claims. Skill self-claims are the most basic form of skill endorsements. Persons can use them to declare

[1] Distributed profiles are documents, which are accessible from different URLs and hosted on different servers, referencing each other. They describe persons.

that they have a certain skill set. So, Alice can claim she has a specific skill like C# programming. To increase the credibility of self-claimed skills, other persons should be enabled to affirm them. Skill affirmations allow persons to testify that someone they know has a specific skill. This affirmation may be based on past collaboration where certain skills were involved and demonstrated. For example, Bob can affirm Alice's C# skill because he worked with her on a project requiring this particular skill. In endorsing someone's skill, the endorser uses his own reputation to give more weight to the endorsee's claimed skill. This contributes to increasing the credibility of claimed skills. Achieving this objective requires delivering a procedure that allows persons to affirm self-claimed skills.

Objective 3: To allow assigning skills the endorsee did not consider. In many respects skill assignments are similar to skill affirmations. A skill assignment suggests a person, like Alice, to claim a skill she has not considered so far. As an example, Bob knows Alice very well. So, he might assign Alice the HTML skill she did not think of. While skill affirmations rely on prior skill self-claims, skill assignments do not. For achieving this objective, we need to deliver a procedure that allows persons to assign skills that have not been self-claimed beforehand.

Objective 4: To prevent unwanted skill endorsements. Centralized social networks can easily incorporate the concept of skill affirmations and skill assignments. They form a single point of truth. The skill endorsements are part of the database of the networking platform. Unless integrity of the data stock has been violated, it is impossible for Mallory to claim negative skills upon Alice. That is, the endorsee needs to self-claim skills beforehand or confirm an assignment.

Adopting this policy to distributed social networks without a central data base is more complicated. First, we must avoid maliciously negative affirmations and assignments. Otherwise, Mallory could affirm negative skills to damage the Alice's reputation by publicly claiming Alice has an "incompetence" skill. Second, persons might be found by expert finding systems due to outdated affirmations of skills they deliberately removed from their profiles. For example, an engineer who has been working for arms industry but now decided against this branch removes corresponding skills from his profile. Distributed expert finding should not consider outdated skill endorsements. Therefore, we need to strive for an agreement between the endorser and the endorsee. As a side effect, this would also contribute to increasing the credibility of skill claims.

Objective 5: To express skill endorsements in distributed user profiles. When Alice claims she has a certain skill set, this information must be recognizable by all authorized members of the distributed social network. The same holds true, when Bob endorses a skill of Alice or when Charlie makes a skill assignment. All three kinds of skill endorsements differ in who is claiming which skill for whom. Thus, each skill endorsement consists of three basic elements: endorser, skill and endorsee. So, a vocabulary able to express such triples in a unified and linkable manner would allow covering all kinds of skill endorsements. Associating and storing skill endorsements with the person claiming them, as suggested in Objective 1, requires delivering a vocabulary for specifying skill endorsements.

Objective 6: To incorporate skill endorsements in distributed expert finding. Achieving Objective 1 fulfills the basic requirements to incorporate skill endorsements in distributed expert finding. The expert finder needs to compare all skill endorsements associated to a candidate with the skills required for a task. For determining a person's suitability for a case, distributed expert finding must consider all kinds of skill endorsements. This assists Casey in deciding about assigning a task to Alice, Bob etc.. To achieve this objective, we need to deliver 1) a method to compare skill endorsements with case requirements and 2) a ranked list of experts fitting to the case requirements.

Objective 7: To facilitate a differentiated consideration of skill endorsements. Case owners benefit from an extensive knowledge about a candidate's suitability for a case. Taking all kinds of skill endorsements into account would enable Casey to gain a rich picture of each candidate's capabilities. Depending on the quantity and quality of a personal social network, the number of skill endorsements differs from person to person. For example, Alice's many social connections also entail many skill affirmations and assignments. The number of skill endorsements could be one criterion for Casey. She knows, however, that this would discriminate persons who have fewer or less diligent social connections.

Distributed expert finding could address such issues by statically weighting each kind of skill endorsement differently. This would, however, reduce adaptability of expert finding and favor persons who share similar characteristics. To preserve customizability, distributed expert finding has to enable adaptably factoring in all kinds of skill endorsements. So, Casey could weight skill self-claims more than skill affirmations or assignments. This is in line with Objectives 2 to 4.

3 Expert Finding with CRAWL

In this section we describe how CRAWL [15] assists expert finding in distributed social networks. The scenario shown in Figure 3 demonstrates our approach. Casey works as a second-level-support worker for a software development company. A key customer reports a bug in a software product developed by the company. Casey is responsible for the handling of this support case. She uses an ACM system to assist her work. As she investigates the problem, she defines several goals and asks experts from the third-level-support department to contribute. At some point during the analysis of the bug, a detailed profiling is required to rule out concurrency issues. However, there is no expert on this topic available. To assist Casey in finding a person with the required expertise, CRAWL facilitates the following workflow (cf. numbers in Figure 3):

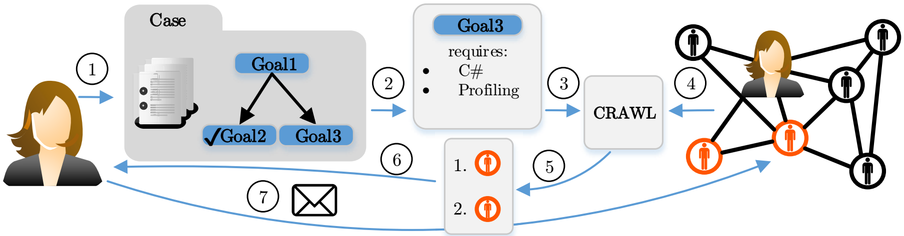

Fig. 3. CRAWL Overview

1. Casey adds a corresponding goal to the case.
2. Casey defines requirements (e.g., C# and Profiling).
3. Casey starts CRAWL.
4. CRAWL traverses Casey's social graph.
5. CRAWL generates a list of eligible workers.
6. Casey selects the most suitable candidates.
7. Casey asks them for contribution to the goal.

Finding suitable workers requires a traversal of the requestor's social graph. This graph is established by `foaf:knows` connections in WebID profiles. WebID profiles are essential artifacts of the WebID identification approach. They contain an identity owner's personal data described in a machine-readable way using Linked Data. For this, WebID relies on several RDF-vocabularies such as FOAF. With WebID, users are enabled to globally authenticate themselves, connect to each other, manage their profile data at a self-defined place and specify customized views [23]. Users can rely on WebID identity providers for creating new WebID identities and managing their WebID profile data [24].

The traversal algorithm is implemented as a depth-limited breadth-first search. It dequeues a WebID URI identifying a person, retrieves the corresponding WebID profile, calculates the rating R, marks the WebID URI as visited and adds all unvisited WebID URIs referenced via `foaf:knows` and their depth value to the queue. The initial queue consists of the WebID URIs of the persons already involved in the case. A maximum depth is used due to the exponentially rising number of nodes in a social graph with increasing depth [13]. CRAWL allows for additional limits like the number of suitable candidates rated above a certain threshold. Following the rating, the WebID profile graph of the candidates is added to a triplestore. A statement containing the calculated rating is asserted into the graph. The final ordered list of rated candidates results from executing the SPARQL query shown in Listing 1.1 on the triplestore.

Sequential traversal of WebID profiles and rating calculation have a huge impact on performance due to the distributed nature of profiles. We addressed this issue by concurrency and caching of user profiles and skill descriptions [15].

```
SELECT ?candidate ?rating
WHERE { ?candidate a foaf:Person .
  ?candidate vsrcm:rating ?rating.
  FILTER( ?rating > ?minRating )}
ORDER BY DESC( ?rating )
```

Listing 1.1. SPARQL query for candidates

Having retrieved and rated a subset of the social graph, CRAWL presents a list of recommended candidates and contact information to the person initiating the search. This step allows for later extension to enable applying constraint criteria, e.g., filter candidates from a specific company or within the same country. CRAWL demonstrates the basic concept of expert finding in distributed social networks leveraging knowledge from profiles, case descriptions and Linked Open Data (LOD) [15]. Therefore it addresses Objective 1. Yet, it does not consider knowledge of people about people such as skill assignments and skill affirmations.

4 CRAWL·E: Extending CRAWL with Endorsements

In order to addresses all objectives from Section 1, we propose CRAWL·E which extends CRAWL with endorsements. The first part of this section introduces a vocabulary to express skill endorsements. Part two explains the expert finding algorithm and the integration of endorsements in the candidate rating.

4.1 Integrating Skill Endorsements in Distributed Profiles

In [15] we introduced a vocabulary to add skill self-claims to WebID profiles. Linked Data provides CRAWL with a large knowledge base for concepts describing skills. CRAWL references this data to describe existing experience for persons and experience required to achieve a case goal or contribute to it. In a WebID profile, the RDF property `vsrcm:experiencedIn` connects a `foaf:Person` with a URI which represents this person's experience in *something*. For referring to the actual skills URIs are used to reference concepts which are available as dbpedia[2] resources. With dbpedia being a central element of the linked open data cloud, this intends to increase the degree of reusability and extensibility of skill data.

To express endorsements, we reuse this vocabulary as seen in Listing 1.2. The important aspect to note is the distributed nature of profiles. An endorser has no write access to foreign endorsees' profiles.

As there is no specific platform or protocol defined for adding statements to WebID profiles, skill assignments have to be expressed in the endorser's own profile. Leveraging the RDF data model and FOAF vocabulary, CRAWL·E enables persons to add skill assignments to their WebID profiles. These skill assignments

[2] `http://dbpedia.org/`

```
@prefix endorser: <http://company.org/>.
@prefix endorsee: <http://minisoft.ru/>.

<endorser:bob> a foaf:PersonalProfileDocument;
    foaf:primaryTopic <endorser:bob#me>;
    foaf:title "Bob Endorser's WebID profile".

<endorser:bob#me> a foaf:Person;
    foaf:name "Bob Endorser";
    foaf:knows <endorsee:alice#me>;
    cert:key [a cert:RSAPublicKey ;
              cert:exponent 65537   ;
              cert:modulus "1234..."^^xsd:hexBinary].

<endorsee:alice#me> vsrcm:experiencedIn <dbp:Linux>,
                                        <dbp:Mysql>.
```

Listing 1.2. Skill assignment in endorser's WebID profile

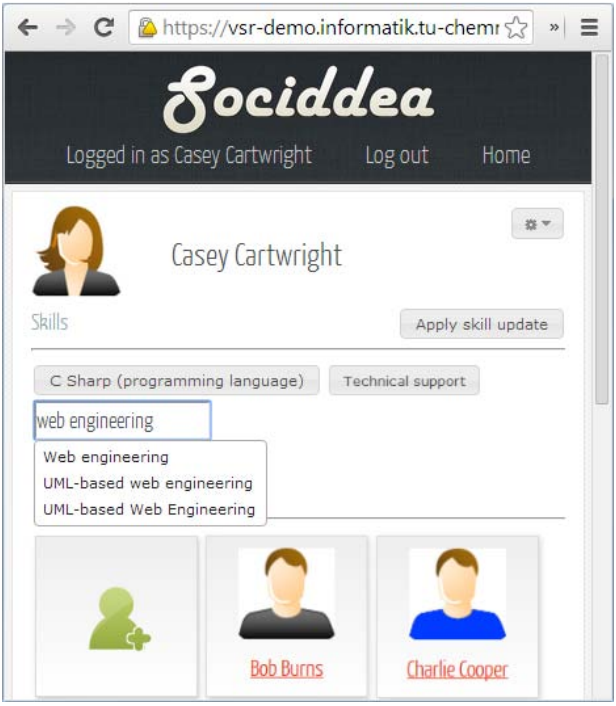

Fig. 4. Skill definition in Sociddea

reference the WebID URI of the endorsee who is connected to the endorser via `foaf:knows`.

Supporting users in specifying their expertise and case requirements, we exemplarily extended the user interfaces of Sociddea and VSRCM [15] to allow specifying skills using regular English words. We use prefix search of dbpedia lookup service to match user input against dbpedia resources. A list of skills is updated live as the user is typing. This is illustrated in Figure 4.

4.2 Extending Distributed Expert Finding to Leverage Skill Endorsements

This section describes how CRAWL·E incorporates endorsements in candidate rating. Figure 5 shows the traversal, rating and candidate recommendation which form steps 4 and 5 in Figure 3. The required skills s_{r0}, s_{r1}, s_{r2} of Goal3 and Casey's social graph are the input. In this example, Casey knows B and C. B and C know D, C knows E. CRAWL·E has already rated B with $R(B) = 15$, C with $R(C) = 0$ and D with $R(D) = 10$. To get the rating of E, the similarities between required skills and existing skills are calculated using linked open data. For a proof-of-concept, we use a prototypical rating function adapted from [16].

According to Objective 5, a skill endorsement is a triple (p_1, p_2, s) of endorser p_1, endorsee p_2 and skill s. A self-claimed skill can be represented as (p, p, s): by an endorsement with identical endorser and endorsee. A candidate c is described by E, the set of all endorsements regarding c in c's social graph as in Equation (1).

$$E = \{(p_1, p_2, s) | p_2 = c\} \tag{1}$$

S_S is the set of self-claimed skills by the candidate, S_O is the set of skills endorsed (assigned) by others and S_B is the set of skills claimed by the candidate and affirmed by others (both) (also cf. to Figure 1).

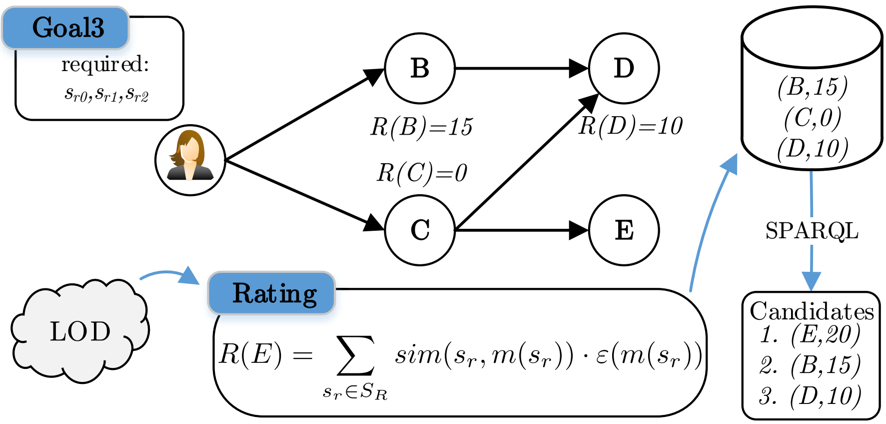

Fig. 5. Traversal, rating and recommendation

$$S_S = \{s|\exists(c, c, s) \in E\} \tag{2a}$$
$$S_O = \{s|\exists(p_1, c, s) \in E, p_1 \neq c\} \tag{2b}$$
$$S_B = S_S \cap S_O \tag{2c}$$
$$S = \{s|\exists(p_1, c, s) \in E\} = S_S \cup S_O \tag{2d}$$

CRAWL·E compares the set of required skills S_R to the set of skills S of each candidate as defined in Equation (2d). Both skill sets are represented by sets of dbpedia URIs. The similarity $sim(s_1, s_2)$ between two skills distinguishes different concept matches:

1. Exact Concept Match - URIs are identical ($s_1 = s_2$)
2. Same Concept As Match - URIs connected via `owl:sameAs` (s_1 `owl:sameAs` s_2)
3. Related Concept Match - URIs connected via `dbprop: paradigm`, `dcterms:subject`, `skos:narrower` etc.

This forms the similarity function in Equation (3).

$$sim(s_1, s_2) = \begin{cases} \sigma_1 & \text{if Exact Concept Match} \\ \sigma_2 & \text{if Same Concept As Match} \\ \sigma_3 & \text{if Related Concept Match} \\ 0 & \text{else} \end{cases} \tag{3}$$

These concept match types can easily be extended to facilitate an adapted rating. The basic idea is that each type yields a different similarity rating. For the moment, we use these values: $\sigma_1 = 10$ for 1), $\sigma_2 = 9$ for 2) and $\sigma_3 = 5$ for 3).

Per candidate, for each combination of a required skill $s_r \in S_R$ and a candidate skill $s_c \in S$, the similarity $sim(s_r, s_c)$ is computed according to Equation (3). As seen in Equation (5), to calculate the candidate rating R in CRAWL, only the skill with maximum similarity per required skill (from function $m : S_R \to S$ in eq. (4)) is considered. [15]

$$m(s_r) = s_c \Leftrightarrow sim(s_r, s_c) = \max_{s \in S} sim(s_r, s) \tag{4}$$

$$R_{\texttt{CRAWL}}(c) = \sum_{s_r \in S_R} sim(s_r, m(s_r)) \tag{5}$$

In CRAWL·E, an affirmed skill is given higher influence compared to a self-claimed skill. To accomodate this influence, we introduce the endorsement factor ε as defined in Equation (7). Our updated CRAWL·E rating function is shown in Equation (6).

$$R_{\texttt{CRAWL}\cdot\texttt{E}}(c) = \sum_{s_r \in S_R} sim(s_r, m(s_r)) \cdot \varepsilon(m(s_r)) \tag{6}$$

Let c be a candidate with the set of endorsements E. The set of all endorsements of the candidate skill s_c is defined by $E_{s_c} = \{(p_1, p_2, s) \in E | s = s_c\} \subseteq E$. With this, we define the endorsement factor using the skill sets defined in 2

$$\varepsilon(s_c) = \begin{cases} 1 & \text{if } s_c \in S_S \setminus S_B \\ \alpha\sqrt{|E_{s_c}|} & \text{if } s_c \in S_B \\ \beta\sqrt{|E_{s_c}|} & \text{if } s_c \in S_0 \setminus S_B \end{cases} \tag{7}$$

This factor distinguishes between the three types of skills: Self-claimed-only skills - from $S_S \setminus S_B$, skills that have been claimed by the candidate and endorsed by others - from S_B - and skills that have only been endorsed by others but are not stated in the candidate's profile - from $S_O \setminus S_B$. It yields 1 for a candidate skill without endorsements, i.e., no additional influence is given to self-acclaimed skills. Parameters α and β allow for adaption, currently we use $\alpha = 1.5$ and $\beta = 2$. To ignore unilateral skill assignments one can set $\beta \stackrel{!}{=} 0$.

The endorsement factor ε increases with the number of endorsements. However, the higher the number of endorsements, the slower the factor increases. This is to avoid overrating candidates with very high endorsement counts. Other function types such as a mirrored $1/x$ function are possible, too. We decided in favor of the square root function type, because it does not converge against a limit as there is no theoretical foundation to reason the limit. When the rating is finished, recommended candidates can be listed as in Figure 6.

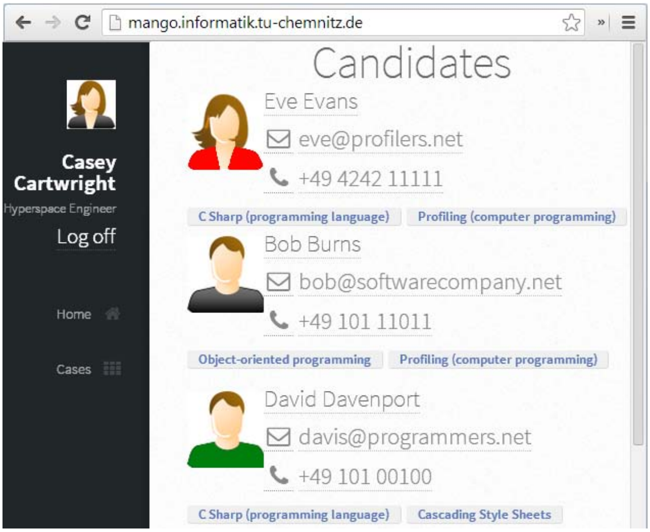

Fig. 6. Candidate recommendation in VSRCM

5 Evaluation

This section discusses the evaluation of our approach. We claim that CRAWL·E achieved the overall objective stated in Section 1 by considering all three kinds of skill endorsements for expert finding in distributed social networks. To prove this claim, we first outline the evaluation setup and then discuss our findings.

5.1 Evaluation Setup

To evaluate the extent to which CRAWL·E's objectives have been achieved, we chose the objective-based evaluation method. Event though a field experiment or a case study would allow for a profound review also, CRAWL·E's results highly depend on the underlying distributed social network. There are various characteristics to be considered, including total and average amount of social connections and of each kind of skill endorsements, richness of user profile data, and level of networking. While a prototypical implementation of CRAWL·E is publicly available, the adoption[3] by users has not yet reached a certain level. Such adoption is, however, required for conducting a field experiment and gaining both extensive and reliable evaluation results. Yet, we are convinced that enabling knowledge workers to find the right collaborators impacts the performance of future enterprises [5]. Thus, the value add by skill endorsements will soon trigger a demand to incorporate them into distributed worker profiles and expert finding.

An objective-based study is the most prevalent approach in program evaluation and is applicable to be performed internally by program developers [20]. As Stufflebeam describes, the "objectives-based approach is especially applicable in assessing tightly focused projects that have clear, supportable objectives" and that "such studies can be strengthened by judging project objectives against the intended beneficiaries' assessed needs, searching for side effects, and studying the process as well as the outcomes". For devising CRAWL we already defined objectives in Section 2. They are well-suited to be reused as evaluation criteria in this objective-based study. We incorporate the information collected during development and application of CRAWL and CRAWL·E, cf. Heil et al. in [15], to determine how well each operational objective was achieved. After outlining the evaluation setup, we discuss the findings for each objective in the following.

5.2 Discussion of Findings

To enable expert finding in distributed social networks, we follow the idea of finding suitable experts to invite. We think that searching for experts by utilizing personal social graphs is more beneficial compared to the open call approach discussed in Crowdsourcing research. Case owners intending to delegate tasks know their social network. So, they know whether a candidate fits the task description. CRAWL·E allows describing skills associated persons and requirements associated to cases. Rather than building our own database to manage skills and skill descriptions as typical for centralized social networks, our approach relies on the collective knowledge of dbpedia. In CRAWL·E, linking to a dbpedia resource refers either to a skill or to a requirement. Unlike related work, we do not want to restrict the number of options for skills and requirements. The availability of a description for a referred concept is not just a requirement in our approach, but also a good practice in general. dbpedia is a central part of the Linked

[3] With regard to the quantity of skill self-claims, skill affirmations and skill assignments per user and in total, and the use of the skill endorsement vocabulary in general.

Open Data cloud. So, resources are machine-readable through RDF and highly connected to each other. This allows for classification and association. The set of resources as well as each resource as such is extensible and maintained by a large community. When attaching skills to persons, we benefit from skills that are both referenced and identified by URIs. Similar to a skill URI pointing to a skill description, we use a WebID URI to refer to a person. Contrary to creating separate resources for storing each person's skill endorsements, we embedded them in machine-readable WebID profiles. As all skill endorsements a person, like Bob, issues belong to him, they also remain in his profile. With all concepts described using RDF, discovery and query become possible through interlinkage with URIs and retrieval using SPARQL. Thus, we delivered all four results necessary for achieving this objective.

To increase credibility of skill self-claims, we introduced the concept of skill affirmations. With a skill self-claim, a person describes that he possesses this skill. The credibility of skills self-claimed by a person, however, depends on the level of trust in this person. A skill self-claim which has been endorsed by someone else gains credibility depending on the trust in the endorsing person. In CRAWL·E, we assume that the more persons have endorsed a person for a particular skill, the more likely is this person to *really* possess the skill. Thus, the higher endorsed a certain skill self-claim is, the more influence it has on the candidate ranking.

To allow assigning skills the endorsee did not consider, our approach enables endorsers to claim skills a person has, without the person having to self-claim those skills beforehand. That is, it is not required for an endorsee to state such skills in their own WebID profiles. This is useful for instance to provide a more complete skill profile that includes information which the described person did not think of. Skill assignments available in distributed social networks can be exploited for various purposes including requests for adding assigned skills to own user profiles. While the increased expressiveness by skill assignments allowed achieving this objective, it comes at the cost of loosing control about what is being endorsed. We addressed this issue as explained in the following paragraph.

To prevent unwanted skill endorsements, we imposed the requirement that endorser and endorsee are bilaterally connected. This indicates that both persons *deliberately* know each other and, hence, accept each other's opinion. As an example, Charlie assigned a certain skill, like COBOL programming, to Alice some time ago. She followed Charlie's suggestion and claims this skill herself. So, Charlie's assignment became an affirmation. Alice was also endorsed by other persons for this skill, i.e., Bob or even Mallory. Today, she is not that interested in this topic anymore. Therefore, Alice does not want to be found for this skill. She removes her self-claimed COBOL skill. Thus, all affirmations become assignments. By excluding Mallory from her social network, all his affirmations and assignments are going to be ignored in CRAWL·E. That is, affirming or assigning a skill necessitates that both the endorsee and the endorser know each other. In addition to this approach, we enable avoiding malicious and outdated skill endorsements by

1) simply removing corresponding skill self-claims the endorsee made in his profile and 2) appropriate consideration within distributed expert finding.

To express skill endorsements in distributed user profiles, we had to find an alternative to what is known from centralized social networks. There, Bob can endorse Alice for a skill she did not think of and the platform triggers a notification to Alice. The endorsed skill will not be added to her profile unless she approves of it. This is not possible with distributed social networks. Alice' and Bob's WebID profiles are documents typically hosted on different servers, and identified and retrievable by URIs. There is no platform to trigger a notification to Alice let alone to allow Bob writing to her profile. Objectives 2 and 3 are inherently adverse to Objective 4 because there is no standard means of asking approval. We therefore delivered an RDF vocabulary to specify skill endorsements using only one RDF triple per skill definition, no matter if it is a skill self-claim, affirmation or assignment. Associating endorser, endorsee and endorsed skill, and storing the triple within the endorser's WebID profile allows for expressing all three kinds of skill endorsements. Due to RDF's flexible and extensible nature, skill endorsements can be attached to the endorser (for self-claims) or to one of his social connections expressed via `foaf:knows` (for affirmations and assignments).

To incorporate skill endorsements in distributed expert finding, CRAWL·E queries personal social networks for skill information, retrieves and processes the skill endorsements, and compares them with the case requirements before showing them a ranked list of suited candidates. Our approach hereby involves how we have addressed Objective 5 for query and retrieval, and Objectives 2 to 4 for comparing and ranking. As manual assessment by crawling personal social networks is time-consuming, our approach assists case owners in their recruitment tasks, but also leaves the final assessment decision to them. While CRAWL only considers self-claimed skills for the candidate recommendation, CRAWL·E also takes skill affirmations and assignments into account. However, all skill endorsements made by persons not knowing each other are ignored in this process. For comparing case requirements with all three kinds of skill endorsements, our approach computes the similarity between both concepts. Although CRAWL·E differentiates exact from similar or related concepts via different weights, finding precise and profound weights requires further empirical evaluation.

To facilitate a differentiated consideration of skill endorsements, we developed CRAWL·E in a parametrized way which allows to choose whether to include foreign endorsed skills or to operate on confirmationary endorsements only. By employing our approach, case owners can adjust the influence of unilaterally endorsed statements, i.e., which value a statement is given that potentially many others claim upon a person but which this person does not claim himself. To reduce the effect of unusually many skill endorsements per person through a large amount of diligent social connections, we introduced a function for only partially factoring in large accumulations of skill affirmations and assignments. Considering Equation (7), CRAWL·E facilitates fine-tuning and even excluding the impact skill affirmations and skill assignments have on distributed expert finding. Users can align the rating with their individual needs and preferences. Finding a more exact function type and evidenced-based default values for the parameters requires a larger empirical evaluation to be conduct in future work.

6 Related Work

Expert Finding has long been a research interest. For example, Becerra-Fernandez provides an overview on Web-based expert finding approaches in [3]. Like [2], many of them are based on information retrieval techniques. To achieve expert finding, they analyze document topics and connect them to the document authors. Perugini et al. surveys expert finding systems with a focus on social context including communication and blogs [19]. Particularly for the research domain, there are several approaches, e.g., by Xu et al. [25] or by Uddin et al. [22].

The approach described by Xu et al. in [25] is similar to CRAWL·E in that it unites social graphs with skill relationship semantics. To achieve this, network analysis on interlinked concept (expertise) and research (social) layer is employed. While this approach also considers hierarchical and correlation relationships in the expertise layer, tacit knowledge is used. CRAWL·E uses explicit knowledge from profiles and Linked Open Data, whereas [25, 2] extract information from unstructured text sources, [25] supported by WordNet. This works well in a specific domain like research, because characteristics of the domain can be exploited. For instance, citations and co-authorship can be analysed from publications [25].

By contrast, CRAWL·E is a generic approach not limited to a specific domain. None of the above approaches works in distributed social networks, nor do they explicitly consider endorsements.

In [6], Bozzon et al. present an expert finding method based on user's activities in centralized social networks like Facebook, Twitter etc. It analyzes social resources directly related (e.g. tweets, likes) or indirectly related (e.g. posts of liked pages) to persons. This approach employs text analysis: entity recognition for skills is performed on the resources, they are identified with Wikipedia URIs. CRAWL·E by contrast targets distributed social networks, uses explicit expertise information, supports skill endorsements and leverages linked data.

In spite of their benefits, skill endorsements have not yet gained much attention in research. Platforms like LinkedIn[4] and ResearchGate[5] have successfully included them. Within the first six month, more than 2 billion endorsements were created on LinkedIn allowing for interesting analysis [14]. Donston-Miller states in [10] that endorsements provide a streamlined version of a resume and can reduce the risk of hiring new personnel. Also quality aspects should be considered in addition to mere quantity (endorsement count) measures. Doyle also mentions in [11] the problem of unwanted endorsements and argues that getting the "right" endorsements is important. While Berk suspects that LinkedIn is using endorsement data in its secret search algorithm [4], there is not much public

[4] http://www.linkedin.com/

[5] http://www.researchgate.net/

information available about skill endorsements in expert finding. Even if current platforms are internally implementing this, the major difference to CRAWL·E is the central nature of these social networks.

Pérez-Rosés et al. presents in [18] an approach which combines social graphs with skill endorsements. It uses an undirected social graph and a directed endorsement graph per skill. Skill relationships like correlation or implied skills are considered. The PageRank algorithm is applied to a deduced graph. Its deduction matrix is similar to the similarity matrix in CRAWL·E. However, the definition of the deduction matrix is an open problem whereas we get the values of the matrix leveraging Linked Open Data. Friendship-like bilateral relationships between social network members are assumed, while the `foaf:knows` semantics employed in CRAWL·E allows for unilateral relationships. Unlike CRAWL·E, this work focuses on social graph analysis, lacks a complete expert finding workflow, and does not support distributed social networks.

Our approach is an application of the *social routing principle* [12] to the ACM domain. Unlike task delegation through an open call known from Crowdsourcing research [7], we follow the idea of inviting suitable experts to contribute to a case by utilizing social graphs. The *conceptual routing table* described by Dustdar and Gaedke is formed by `foaf:knows` statements and contact info in WebID profiles.

7 Conclusions and Future Work

In this work, we presented the CRAWL·E approach leveraging distributed endorsements for finding eligible workers to contribute to ACM cases. It comprises a vocabulary for skill endorsements in WebID profiles, a method for traversing distributed social networks based on foaf:knows relationships and an adaptable rating function for WebID profiles. We demonstrated CRAWL by implementation based on the WebID identity provider and management platform Sociddea and the case management system VSRCM.

Our future research interest will be to consider not only endorsement quantity, but also quality. If a renowned expert endorses someone else for his very own field, his endorsement should be given more weight compared to endorsements of less renowned persons. This needs considering the endorsements of the endorsers in addition to the endorsements of the candidate to rate. Empirical data and machine learning can be used to provide adapted parameters. Providing Distributed Expert Finding as a Service is desirable to enable easy integration in other systems. For this, an endpoint structure and protocol must be defined.

References

[1] Appelquist, D., et al.: A Standards-based, Open and Privacy-aware Social Web: W3C Incubator Group Report. Tech. rep., W3C (2010)
[2] Balog, K., Azzopardi, L., de Rijke, M.: Formal models for expert finding in enterprise corpora. In: Proceedings of the 29th Annual International ACM SIGIR Conference, pp. 43–50. ACM, New York (2006)

[3] Becerra-Fernandez, I.: Searching for experts on the Web: A review of contemporary expertise locator systems. ACM TOIT 6(4), 333–355 (2006)
[4] Berk, R.A.: Linkedin Triology: Part 3. Top 20 Sources for Connections and How to Add Recommendations. The Journal of Faculty Development 28(2), 1–13 (2014)
[5] Bider, I., Johannesson, P., Perjons, E.: Do workflow-based systems satisfy the demands of the agile enterprise of the future? In: La Rosa, M., Soffer, P. (eds.) BPM Workshops 2012. LNBIP, vol. 132, pp. 59–64. Springer, Heidelberg (2013)
[6] Bozzon, A.,et al.: Choosing the Right Crowd: Expert Finding in Social Networks Categories and Subject Descriptors. In: Proceedings of the 16th International Conference on Extending Database Technology, New York, NY, USA, pp. 637–348 (2013)
[7] Brabham, D.C.: Crowdsourcing as a Model for Problem Solving: An Introduction and Cases. Convergence: The International Journal of Research into New Media Technologies 14(1), 75–90 (2008)
[8] Clair, C.L., Miers, D.: The Forrester Wave™: Dynamic Case Management, Q1 2011. Tech. rep., Forrester Research (2011)
[9] Davenport, T.H.: Rethinking knowledge work: A strategic approach. McKinsey Quarterly (2011)
[10] Donston-Miller, D.: What LinkedIn Endorsements Mean To You (2012), `http://www.informationweek.com/infrastructure/networking/what-linkedin-endorsements-mean-to-you/d/d-id/1106795`
[11] Doyle, A.: How To Use LinkedIn Endorsements (2012), `http://jobsearch.about.com/od/linkedin/qt/linkedin-endorsements.htm`
[12] Dustdar, S., Gaedke, M.: The social routing principle. IEEE Internet Computing 15(4), 80–83 (2011)
[13] Goel, S., Muhamad, R., Watts, D.: Social search in “small-world” experiments. In: Proceedings of the 18th International Conference on World Wide Web, WWW 2009, pp. 701–710. ACM, New York (2009)
[14] Gupta, S.: Geographic trends in skills using LinkedIn’s Endorsement feature (2013), `http://engineering.linkedin.com/endorsements/geographic-trends-skills-using-linkedins-endorsement-feature`
[15] Heil, S., et al.: Collaborative Adaptive Case Management with Linked Data. To appear in WWW 2014 Companion: Proceedings of the 23rd International Conference on World Wide Web Companion, Seoul, Korea (2014)
[16] Lv, H., Zhu, B.: Skill ontology-based semantic model and its matching algorithm. In: CAIDCD 2006, pp. 1–4. IEEE (2006)
[17] Mundbrod, N., Kolb, J., Reichert, M.: Towards a system support of collaborative knowledge work. In: La Rosa, M., Soffer, P. (eds.) BPM Workshops 2012. LNBIP, vol. 132, pp. 31–42. Springer, Heidelberg (2013)
[18] Pérez-Rosés, H., Sebé, F., Ribó, J.M.: Endorsement Deduction and Ranking in Social Networks. In: 7th GraphMasters Workshop, Lleida, Spain (2013)
[19] Perugini, S., Goncalves, M.A., Fox, E.A.: A connection-centric survey of recommender systems research. Journal of Intelligent Information Systems 23(2), 107–143 (2004)
[20] Stufflebeam, D.: Evaluation Models. New Directions for Evaluation 2001(89), 7–98 (2001)
[21] Swenson, K.D.: Position: BPMN Is Incompatible with ACM. In: La Rosa, M., Soffer, P. (eds.) BPM Workshops 2012. LNBIP, vol. 132, pp. 55–58. Springer, Heidelberg (2013)
[22] Uddin, M.N., Duong, T.H., Oh, K.-j., Jo, G.-S.: An ontology based model for experts search and ranking. In: Nguyen, N.T., Kim, C.-G., Janiak, A. (eds.) ACIIDS 2011, Part II. LNCS, vol. 6592, pp. 150–160. Springer, Heidelberg (2011)

[23] Wild, S., Chudnovskyy, O., Heil, S., Gaedke, M.: Customized Views on Profiles in WebID-Based Distributed Social Networks. In: Daniel, F., Dolog, P., Li, Q. (eds.) ICWE 2013. LNCS, vol. 7977, pp. 498–501. Springer, Heidelberg (2013)
[24] Wild, S., Chudnovskyy, O., Heil, S., Gaedke, M.: Protecting User Profile Data in WebID-Based Social Networks Through Fine-Grained Filtering. In: Sheng, Q.Z., Kjeldskov, J. (eds.) ICWE Workshops 2013. LNCS, vol. 8295, pp. 269–280. Springer, Heidelberg (2013)
[25] Xu, Y., et al.: Combining social network and semantic concept analysis for personalized academic researcher recommendation. Decision Support Systems 54(1), 564–573 (2012)