

# Expertise bewerben und finden im Social Semantic Web

Axel Polleres<sup>1</sup> and Malgorzata Mochol<sup>2</sup>

<sup>1</sup>Digital Enterprise Research Institute National University of Ireland  
IDA Business Park, Lower Dangan Galway, Ireland  
E-Mail: axel@polleres.net

<sup>2</sup>Freie Universität Berlin, AG Netzbaasierte Informationssysteme  
Takustr. 9, 14195 Berlin, Deutschland  
E-Mail: mochol@inf.fu-berlin.de

**Zusammenfassung** Im vorliegenden Beitrag diskutieren wir Rahmenbedingungen zur Kombination, Wiederverwendung und Erweiterung bestehender RDF Vokabulare im Social Semantic Web. Hierbei konzentrieren wir uns auf das Anwendungsszenario des Auffindens und Bewerbens von Experten im Web oder Intranet. Wir präsentieren, wie RDF Vokabulare mit zunehmendem Verbreitungsgrad im Semantic Web einerseits und de facto Standardformate, die von täglich verwendeten Applikationen benutzt werden, andererseits (z.B. vCard, iCal oder Dublin Core) kombiniert werden können, um konkrete Anwendungsfälle der Expertensuche und zum Management von Expertise zu lösen. Unser Fokus liegt darauf aufzuzeigen, dass für praktische Anwendungsszenarien nicht notwendigerweise neue Ontologien entwickelt werden müssen, sondern der Schlüssel vielmehr in der Integration von bestehenden, weit verbreiteten, und sich ergänzenden Formaten zu einem kohärenten Netzwerk von Ontologien liegt. Dieser Ansatz garantiert sowohl direkte Anwendbarkeit als auch niedrige Einstiegsbarrieren für Semantic Web Technologien, sowie einfache Integrierbarkeit in bestehende Applikationen. Die im Web verfügbaren und verwendeten RDF Formate decken zwar einen großen Bereich der Aspekte zur Beschreibung von Personen und Expertisen ab, zeigen aber auch signifikante Überlappungen. Bisher gibt es wenig systematische Ansätze, um diese Vokabulare zu verbinden, sei es in Form von allgemeingültigen Praktiken, die definieren, wann welches Format zu benutzen ist, oder in Form von Regeln, die Überlappungen zwischen einzelnen Formaten formalisieren. Der vorliegende Artikel analysiert, wie bestehende Formate zur Beschreibung von Personen, Organisationen und deren Expertise kombiniert und wo nötig erweitert werden können. Darüber hinaus diskutieren wir Regelsprachen zur Beschreibung von Formatüberlappungen, sowie deren praktische Verwendbarkeit zur Erstellung eines Ontologie-Netzwerks zur Beschreibung von Experten.

## 1 Einführung

Das Semantic Web ist da! Eine stetig wachsende Anzahl von Individuen oder Institutionen beginnt damit, Metadaten in Form von RDF Annotationen auf ihren persönlichen Homepages oder den Webseiten ihrer Organisation zur Verfügung zu stellen. Neben RDF/XML [27] bieten neue Technologien wie Microformate<sup>1</sup> die Möglichkeit Metadaten einfacher als bisher direkt in beliebige HTML oder XHTML Dokumente einzubetten. Die vom World Wide Web Consortium (W3C) kürzlich ins Leben gerufene GRDDL [17] Arbeitsgruppe<sup>2</sup> entwickelt standardisierte Mechanismen, um semantisch reichere Daten in RDF aus (semi-)strukturierten Daten zu extrahieren. Eine weitere W3C Arbeitsgruppe – Semantic Web Best Practices and Deployment<sup>3</sup> – bietet Richtlinien zur Publikation und syntaktischen Kombination von RDF [27, 12] Daten und OWL [37] Ontologien an. Darüber hinaus helfen Browsererweiterungen wie Semantic Radar<sup>4</sup> RDF Annotationen auf Webseiten zu entdecken, und eine Reihe von generischen RDF Browsern wie Tabulator<sup>5</sup> und Suchmaschinen wie SWSE [23] oder Sindice [44] erlauben Anwendern direkt im Semantic Web zu “surfen”.

Nachdem also die Infrastruktur des Semantic Web beachtliche Fortschritte gemacht hat und syntaktische Hindernisse zum Verbinden von Metadaten weitgehend beseitigt scheinen, wollen wir uns hier genauer mit der Frage befassen, welche RDF Vokabulare und Ontologien tatsächlich verwendet werden können, um eine bestimmte Domäne, in unserem Fall Metadaten über Personen und Organisationen, zu beschreiben. Für diesen Zweck widmen wir uns einer Schlüsselapplikation von Semantic Web Technologien, der im allgemeinen viel Potential beigemessen wird und deren tatsächlicher Verwirklichung aber immer noch einige Hürden im Weg zu stehen scheinen: Das Auffinden von Experten (Individuen, Teams, Organisationen) ist nach wie vor eine manuelle mühsame Aufgabe. Obwohl ein Gutteil existierender Webseiten genau darauf abzielt, Expertisen von Personen oder Organisationen zu bewerben, bleiben Suchmöglichkeiten auf Schlagwörter oder Branchenverzeichnisse, die manuell aggregiert wurden, beschränkt. Unsere Annahme ist, dass im Web beworbene Institutionen, Projekte, Personen und Events durch eine Handvoll maschinenlesbarer Formate beschrieben werden können. Dadurch wird das automatische Auffinden von Expertise/Experten in einer bestimmten Domäne oder für eine bestimmte Aufgabe möglich. Um dieses Ziel zu erreichen, und ähnliche Anwendungsszenarien von semantischer Suche im Web zu verwirklichen, lassen sich drei kritische Erfolgskriterien, die sich gegenseitig beeinflussen, festmachen:

- (i) *Gemeinsam verwendete, maschinenlesbare Formate*
- (ii) *Kritische Masse von Anwendern/Nutzern*
- (iii) *Zusätzlich benötigte Technologien*

<sup>1</sup> <http://microformats.org/>

<sup>2</sup> <http://www.w3.org/2001/sw/grddl-wg/>

<sup>3</sup> <http://www.w3.org/2001/sw/BestPractices/>

<sup>4</sup> <http://sioc-project.org/firefox/>

<sup>5</sup> <http://www.w3.org/2005/ajar/tab>

*Geeignete Formate* und Ontologien sind im Web nur dann nützlich, wenn sie von einer *kritischen Masse* von Anwendern und Applikationen unterstützt werden. Die Erfahrung zeigt, dass nicht unbedingt umfangreiche und detaillierte Ontologien den höchsten Verbreitungsgrad erfahren, sondern oft einfache Formate, die auf das wesentliche reduziert sind, wie beispielsweise FOAF [13] oder SIOC [6]. *Zusätzliche Technologien* sind nötig, um diese Formate miteinander zu verbinden. Dies ist insbesondere dann der Fall, wenn man sich noch nicht auf ein Format oder geeignete Algorithmen geeinigt hat. *Zusätzliche Technologien* wie Regelsprachen und Abfragesprachen werden hier beispielsweise benötigt, um Überlappungen und Übersetzungen zwischen verschiedenen Formaten deklarativ zu beschreiben oder um gezielte Anfragen nach bestimmten Expertenprofilen zu stellen.

Wir werden im Folgenden auf jeden dieser Faktoren im einzelnen eingehen (siehe Abschnitt 2) und Implikationen für die Anwendungsdomäne des Auffindens und Bewerbens von Experten im Web erläutern. Anschließend stellen wir in Abschnitt 3 die ExpertFinder Initiative vor, welche sich zum Ziel gesetzt hat, der Verwirklichung der Versprechen des Semantic Web in dieser Anwendungsdomäne anhand von etlichen konkreten Szenarien einen Schritt näherzukommen. In Abschnitt 4 stellen wir Standardformate und Ontologien vor, welche direkt zur Beschreibung von Experten im Web verwendet werden können oder indirekt eine Bedeutung für diese Domäne haben. Weitere relevante Projekte und Anwendungen werden in Abschnitt 5 diskutiert. Abschließende Bemerkungen samt Ausblick finden sich in Abschnitt 6.

## 2 Erfolgskriterien für Semantic Web basierte Expertensuche

In diesem Abschnitt wollen wir im Einzelnen die drei oben genannten Faktoren im Hinblick auf die Anwendungsdomäne web-basierte Expertensuche diskutieren.

### 2.1 *Gemeinsam verwendete, maschinenlesbare Formate*

Für den Moment nehmen wir an, dass syntaktische Fragen bezüglich der Form, in welcher Metadaten publiziert werden sollen um im Web auffindbar zu sein, geklärt sind, d.h. dass:

- RDF als Standardformat zur Publikation von Metadaten dient,
- Informationsanbieter wissen, wie sie semantische Annotationen zu ihren Webseiten hinzufügen können,
- Werkzeuge zur Verfügung stehen, um diese Metadaten auszulesen und zu sammeln.

Zu klären bleibt die Frage nach semantischen Aspekten, d.h. die Frage nach existierenden und verbreiteten Ontologien und RDF Vokabularen, auf die Informationsanbieter zurückgreifen können, um ihre Expertise und andere relevante Aspek-

te präzise und umfassend zu beschreiben. Neben eigentlichen Qualifikationen und Geschäftssektoren sollen diese Vokabulare auch Aspekte beschreiben, die man typischerweise in Organisationsprofilen oder Lebensläufen findet wie Adressen, Publikationen, Projekte und Projektpartner, Referenzen und relevante Kontakte, aber auch Informationen zur Verfügbarkeit, zu sozialen Netzwerken und dergleichen mehr. In Abschnitt 4 werden wir relevante RDF Vokabulare, die all diese Bereiche abdecken, analysieren und sie evtl. vorhandenen Standards aus der Geschäftswelt gegenüberstellen.

## 2.2 *Kritische Masse*

Unabhängig davon, welche(s) RDF Vokabular(e) wir zur Beschreibung von Expertise im (Semantic) Web verwenden wollen – um Webskalierbarkeit zu erreichen, ist es unumgänglich, dass:

- (i) eine kritische Masse von Anwendern und Informationsanbietern überzeugt wird, die vorgeschlagenen Formate zu benutzen oder zu unterstützen,
- (ii) existierende Inhalte in die gewählten Formate übersetzt/importiert werden können,
- (iii) Mechanismen zur Verfügung stehen, um die ausgewählten Formate mit bereits im Web verwendeten Formaten zu kombinieren (z.B. unter Verwendung von `owl:equivalentClass`, `owl:equivalentProperty`) bzw. die existierenden Formate weitestgehend wiederzuverwenden.

In Bezug auf Punkt (i) sei angemerkt, dass eine neue Ontologie zu entwickeln, eine aktive User Community zu akquirieren, bzw. eine neue Ontologie in einer bestehenden Community zu etablieren sehr schwierig ist. In der offenen Community der Webbenutzer kann man annehmen, dass dieses Unterfangen nicht nur wenig erfolgversprechend sondern schlichtweg unmöglich ist. Individuelle User oder Organisationen benutzen Teile von Ontologien, erweitern diese für ihre eigenen Bedürfnisse und benutzen verschiedene URIs um gleiche oder zumindest überlappende Konzepte zu beschreiben. Bezüglich Punkt (ii) haben Textextraktionsmethoden oder Wrapper Technologien<sup>6</sup> in den letzten Jahren mehr Stabilität erreicht. Projekte wie PiggyBank [25] haben gezeigt, dass gemeinsame Nutzung von dezentralen Wrappern ein durchaus erfolgreicher Ansatz sein kann, um RDF Daten aus bestehenden Quellen im Web zu gewinnen. In diesem Zusammenhang könnten Semantic Web Pipes [35] im Stil von Yahoo's Web Pipes<sup>7</sup> der nächste logische Schritt sein. Dennoch lösen solche Ansätze das Problem nur zum Teil, da die Qualität von "Wrappern" oder "Pipes" stark von der Qualität und Stabilität der Formate der Quelldaten

---

<sup>6</sup> Der Überbegriff "Wrapper" wird oft für semi-deklarative Extraktionsmethoden wie XSLT [18] oder spezielle Tools verwendet, die semi-strukturierte Daten wie Webseiten oder allgemeiner XML Formate in andere XML Formate oder auch RDF übersetzen.

<sup>7</sup> <http://pipes.yahoo.com/>

abhängig ist. Das grundlegende Problem bezüglich der Frage, welche RDF Ontologien/Vokabulare für die generierten Metadaten zu verwenden sind, bleibt bestehen. Außerdem gibt es in den meisten Domänen, speziell im Bereich Expertensuche nicht “*die richtige*” Ontologie, weshalb sich unsere Arbeit hauptsächlich auf Punkt (iii) konzentriert. Wir wollen versuchen, so effektiv wie möglich existierende Formate, welche sich durch einen gewissen *Verbreitungsgrad* bereits etabliert haben, wiederzuverwenden. Zu diesem Zweck, wollen wir bestehende Formate analysieren, eventuelle Überlappungen formal beschreiben, und Richtlinien erstellen, wie diese Formate gemeinsam zu verwenden sind. Die ExpertFinder Initiative folgt dieser Idee und in den folgenden Abschnitten werden wir näher auf den derzeitigen Stand unserer Analyse bestehender Vokabulare eingehen. Ein praktischer Nebeneffekt der Wiederverwendung bestehender Formate und Vokabulare ist neben Feedback von bestehenden User Communities, dass wir auf bestehende Tools zurückgreifen können. Formate wie iCal [16], vCard [46], oder BibTeX [36] und deren RDF Varianten werden beispielsweise von Werkzeugen wie Kalender- und Adressbuchapplikationen, oder Online Publikationsverzeichnissen wie DBLP<sup>8</sup> und Citeseer<sup>9</sup> unterstützt, und können somit als Standardaustauschformate für die jeweilige Teildomäne, die sie abdecken, angesehen werden.

### 2.3 Zusätzlich benötigte Technologien

Während wir uns bisher ausschließlich damit befasst haben, wie die nötigen Metadaten im Web repräsentiert werden, wollen wir uns nun fragen, welche zusätzlichen Technologien fehlen, um diese Metadaten für praktische Anwendungsfälle (mehr dazu in Abschnitt 3) zu nutzen. Hier spielen einerseits, Regeln, z.B. in Form von formalisierten Business-Strategien und -Richtlinien, aber auch Empfehlungsalgorithmen, kollaborative Filter-Algorithmen, statistische Methoden, etc. eine wichtige Rolle. All diese Methoden können hilfreich sein um Metadaten zu generieren und zu verbinden, Suchergebnisse richtig zu reihen und zu bewerten. Zudem werden Mechanismen zur Sicherstellung von Datenschutz und Datensicherheit, sowie Verhandlung von Vertrauensverhältnissen zwischen Partnern benötigt, um zu vermeiden, dass personenspezifische oder andere schützenswerte Daten im Web in die falschen Hände geraten. In mehreren dieser Teilfragen, werden Regelsprachen und Regelverarbeitungssysteme eine Schlüsselrolle spielen.

- (i) Solche Regeln (gemeinsam mit ausdrucksstarken Ontologiesprachen wie OWL) erlauben uns die exakten Verbindungen und Überlappungen zwischen bestehenden RDF Vokabularen und Ontologien formal zu beschreiben.
- (ii) Durch Regeln lassen sich Richtlinien ausdrücken, die zum Beispiel festlegen, unter welchen Umständen man bereit ist, bestimmte Metadaten zu teilen oder an dritte weiterzugeben [10].

---

<sup>8</sup> <http://dblp.uni-trier.de/>

<sup>9</sup> <http://citeseer.ist.psu.edu/>

Bezüglich Punkt (i) stellen wir fest, dass einerseits viele RDF Vokabulare wie z.B. FOAF und SIOC zwar schon weitgehend in OWL formalisiert sind, aber viele Features von OWL gar nicht nutzen und größtenteils mit Sprachkonstrukten des ausdrücksschwächeren RDF Schema [12] auskommen. Andererseits jedoch sind die Möglichkeiten von OWL nicht ausreichend, um die Überlappungen zwischen bestehenden RDF Vokabularen exakt zu beschreiben, wie wir an einem kleinen Beispiel zeigen können: Nehmen wir an, wir wollen RDF Daten austauschen, welche einerseits das FOAF [13] Format und andererseits vCard [26] verwenden, d.h. wir wollen von RDF Tripeln, die das `vCard:homeTel` Attribut verwenden nach `foaf:phone` übersetzen, wobei ersteres ein sogenanntes Datentyp-Property (dessen Wertebereich RDF Literale, also Zeichenketten, vorschreibt) und letzteres ein sogenanntes Objekt-Property (dessen Wertebereich RDF Ressourcen, also URIs, vorschreibt) ist. So eine Übersetzung lässt sich klarerweise nicht durch die normale Subproperty-Relation in RDFS oder OWL ausdrücken, da hier eine Wertkonversion nötig ist, um eine URI aus einem RDF Literal zu generieren.

Nachdem das W3C bisweilen noch keinen Standard entwickelt hat, um solche Übersetzungen deklarativ zu beschreiben<sup>10</sup>, könnte man versuchen SPARQL's CONSTRUCT Anfragen zur Beschreibung solcher Übersetzungsregeln zu "missbrauchen". SPARQL [40] ist die kommende Abfragesprache für RDF Daten, welche kurz vor Standardisierung durch das W3C steht. CONSTRUCT-Anfragen in SPARQL dienen dazu, neue RDF Daten aus bestehenden RDF Graphen zu generieren.

Für unsre Beispielübersetzung müssten wir eine Featureerweiterungen in SPARQL, wie die Verwendung von XPath 2.0 Funktionen [31] in CONSTRUCTs, annehmen:

```
CONSTRUCT
{?X foaf:phone xsd:anyURI (fn:concat ("tel:",fn:encode-for-uri (?T)) ) .}
WHERE
{ ?X vCard:tel ?T . }
```

Diese einfache Abfrage ist in der aktuellen Spezifikation von SPARQL jedoch noch nicht möglich. Entsprechende Erweiterungen werden in [3, 39] ausführlich diskutiert.

Neben Übersetzungsregeln, können Regeln, die gemeinsam mit RDF Daten publiziert werden, auch dazu dienen, implizite Metadaten zu beschreiben [29, 38]. Dies würde es ermöglichen, Metadaten von verschiedenen Informationsanbietern über (möglicherweise negative) Abhängigkeiten zu "verlinken".<sup>11</sup> Hier ein paar für die Expertensuche relevante Beispielregeln:

*Alle aufgelisteten Personen unter der URI <http://www.w3.org/People/> sind Experten in [http://en.wikipedia.org/wiki/Semantic\\_Web](http://en.wikipedia.org/wiki/Semantic_Web).<sup>12</sup>*

<sup>10</sup> Man könnte XSLT [18] verwenden, um von einer RDF/XML Datei in eine andere zu übersetzen, aber das birgt einige mögliche Probleme in sich, da die RDF/XML Darstellung ein und desselben RDF Graphen nicht notwendigerweise eindeutig ist.

<sup>11</sup> Siehe dazu auch [21, Abschnitt 2.10].

<sup>12</sup> Wikipedia-Adresslinks sind eine von vielen Möglichkeiten um Themenbereichen von Experten eine eindeutige URI im Semantic Web zuzuordnen.

*Alle aufgelisteten Personen unter der URI <http://www.myExample.org/Contacts.rdf> die NICHT für aufgelistete Firmen unter <http://www.myCompetitors/list.rdf> arbeiten sind meine Freunde.*

*<http://www.w3.org/People/Berners-Lee/card#i> ist der Autor aller Publikationen die unter <http://dblp.uni-trier.de/Berners=Lee:Tim.html> aufgelistet sind.*

Syntaktisch könnte man hier wieder CONSTRUCT Anfragen von SPARQL als View/Link-Definitionen benutzen, aber ein dezidiertes Standard ist derzeit noch nicht verfügbar. Wir erwarten jedoch, dass das W3C, und insbesondere die RDF Data Access (DAWG) Arbeitsgruppe<sup>13</sup>, die an SPARQL arbeitet, sowie die Rule Interchange Format (RIF) Arbeitsgruppe<sup>14</sup> [7], die derzeit an einem Standardaustauschformat für Regeln im Web arbeitet, in absehbarer Zukunft entsprechende Spezifikationen liefern werden.

### 3 Praktische Anwendungsszenarien der ExpertFinder Initiative

Die ExpertFinder<sup>15</sup> Initiative ist eine lose Vereinigung diverser akademischer Partner und Proponenten der FOAF und SIOC Communities. Die Initiative strebt an, RDF Vokabulare und Regelerweiterungen zu einem größeren Ganzen zu vereinen, um persönliche Homepages, Webseiten von Organisationen, Publikationsverzeichnisse oder auch akademische Konferenzwebseiten mit adäquaten Metadaten zu versehen. Dadurch soll die automatisierte Expertensuche im Web ermöglicht werden. Dazu sollen nicht nur bestehende RDF Vokabulare und deren Überlappungen analysiert werden, sondern auch Richtlinien erstellt werden, wie diese Formate gemeinsam zu verwenden sind. Die Initiative wurde 2006 gegründet, um relevante Forschungsprojekte zusammenzubringen und den in Abschnitt 2 genannten Erfolgskriterien näherzukommen. Ein konkretes Ziel der Initiative ist die Identifikation von tatsächlichen Anwendungsszenarien, in denen Mechanismen zur Expertensuche, basierend auf Semantic Web Technologien, angewandt werden können, und für die ein solches Netzwerk von RDF Vokabularen nützlich sein kann. Im Folgenden werden wir diese Szenarien kurz beschreiben.

#### Generierung von Instituts- und persönlichen Webseiten aus RDF Metadaten

RDF Metadaten sind an sich ein geeignetes Datenformat für Content Management Systeme (CMS), welche dynamische Inhalte vom Layout trennen. Obwohl solche Systeme derzeit meist auf traditionellen, relationalen Datenmodellen beruhen, ist zu erwarten, dass rein RDF basierte CMS weit flexiblere Lösungen erlauben. Viele Forscher und Entwickler, die mit Semantic Web Technologien zu tun haben, beginnen derzeit ihre persönlichen Homepages mit dem Friend-of-a-Friend (FOAF) Vokabular zu annotieren. Mithilfe von CMS, die solche RDF Daten direkt integrieren

<sup>13</sup> <http://www.w3.org/2001/sw/DataAccess/>

<sup>14</sup> <http://www.w3.org/2005/rules/>

<sup>15</sup> <http://www.rdfweb.org/topic/ExpertFinder>

können, könnten einerseits Institutsangehörige ihre persönliche Information in ihrer erweiterten FOAF Dateien up-to-date halten; diese würden automatisch mit dem CMS der Organisation synchronisiert. Zum anderen, kann die Organisation Default-Regeln angeben, um evtl. fehlende Daten zu ergänzen. Solche Regeln können auch die Integration von Metadaten von Drittanbietern (siehe dazu die Beispielregeln in Abschnitt 2.3) beinhalten und würden dynamische, dezentral verwaltete Webseiten ermöglichen.

Nehmen wir an, unser Kollege hat keine Zeit oder Lust, eigene Metadaten für die Institutswebseite zur Verfügung zu stellen. In unsrem RDF-basiertem Szenario wäre das kein Problem: Basisdaten können aus der Universitätsdatenbank übernommen werden, eine Publikationsliste kann aus den Daten von DBLP<sup>16</sup> automatisch generiert werden, und so weiter. Sobald die zugrundeliegenden Quellen ihre Daten ihrerseits in RDF zur Verfügung stellen und gemeinsam verwendete Vokabulare unterstützen, kann eine solche Anwendung mithilfe von Regeln (vgl. Abschnitt 2) ohne größere Schwierigkeiten die notwendigen Daten aggregieren. Indem man gemeinsam verwendete RDF Formate direkt unterstützt, wird Export, Abfragen und die Kombination von Daten verschiedener Quellen zum Kinderspiel. Vorläufige Resultate des Gebrauchs von FOAF und anderen Metadaten-Formaten direkt als Grundlage von Homepages wurden beispielsweise im RDFHomepage Projekt realisiert [22]. Auch in Open Source CMS Architekturen wie Drupal<sup>17</sup> ist die direkte Integration von RDF Datenquellen derzeit ein Thema<sup>18</sup>.

### Personalwesen

Mithilfe von geeigneten RDF Vokabularen können Jobsuchende ihr Profil oder ihren detaillierten Lebenslauf in Form von maschinenlesbaren Metadaten im Web publizieren, oder Angestellte in einer Firma können unter Verwendung derselben Vokabulare ein Profil ihrer Fähigkeiten und Erfahrungen im Intranet zur Verfügung stellen. Im Gegenzug können öffentliche oder private Jobvermittlungsagenturen mittels Web-Agenten präferierte Profile im Netz finden um geeignete Kandidaten für eine offene Stelle zu finden, oder ihrerseits die gesuchten Profile im selben Format im Netz zu veröffentlichen. Des Weiteren könnte die Teambildung innerhalb Organisationen aufgrund von im Intranet verfügbarer Profilinformatoren, die über die Fähigkeiten der Mitarbeiter Aufschluss geben, teilweise automatisiert werden. In diesem Fall wird die Auswahl von geeigneten Mitarbeitern, die eine bestimmte Aufgabe kollaborativ lösen sollen, mithilfe von semantischer Suche und Regeln erleichtert. Im Gegensatz zu derzeit verfügbaren, zentralen Jobportalen soll das ExpertFinder Vokabular ermöglichen, solche Szenarien flexibel und weitestgehend dezentral zu lösen.

---

<sup>16</sup> DBLP Daten stehen bereits in Form von RDF zur Verfügung: <http://dblp.13s.de/d2r>.

<sup>17</sup> <http://drupal.org/>

<sup>18</sup> <http://barcelona2007.drupalcon.org/node/633>,  
<http://barcelona2007.drupalcon.org/node/438>,  
[http://drupal.org/project/semantic\\_search](http://drupal.org/project/semantic_search)



### **Semantisch annotierte Wissenschaftsportale**

Die ExpertFinder Idee ließe sich in ähnlicher Weise wie auf Jobportale auch auf öffentliche Wissenschaftsportale, wie zum Beispiel die erfolgreiche CORDIS Plattform der EU<sup>19</sup> übertragen. Semantische Annotation solcher Portale mit dem ExpertFinder Vokabular würde die gezielte Suche bis hinunter zu Informationen zu einzelnen Forschern ermöglichen. Außerdem wäre es so denkbar, dass jede Institution die entsprechenden Metadaten direkt, dezentral veröffentlicht, anstatt eventuelle Änderungen jedesmal über das CORDIS Portal eingeben zu müssen. Ein solchermaßen dezentrales Szenario stellt selbstverständlich erhöhte Anforderungen im Bezug auf die Zertifizierung vertrauenswürdiger Inhalte: Anstatt ein zentrales Portal zu verwalten, könnten sich öffentliche Institutionen wie die EU darauf beschränken, Metadaten von Institutionen entsprechend eigener Richtlinien zu zertifizieren und spezielle Websuchmaschinen für dermaßen zertifizierte Inhalte anbieten.

### **Semantische Auswahl von wissenschaftlichen Gutachtern**

Im akademischen Bereich unterliegen Beiträge für Konferenzen, Zeitschriften oder Projektanträge einem strengen Begutachtungsprozess. Geeignete Gutachter zu finden, die in der Lage sind Beiträge zu teils sehr speziellen Themen zu bewerten, ist oft kein einfaches Unterfangen. Auf der anderen Seite, werden Autoren bei der heutzutage meist elektronischen Einreichung solcher Beiträge bereits jetzt aufgefordert, ihren Beitrag durch Beschlagwortung oder Zuordnung, beispielsweise in eine von der Association for Computing Machinery (ACM) veröffentlichten Kategorien, semantisch vorzuklassifizieren. Anhand dieser Information, der Analyse der zitierten Beiträge oder weiterer Textanalyse, lässt sich ein relativ genaues Profil für geeignete Gutachter ableiten. Aus Online Publikationsverzeichnissen oder Verzeichnissen von Programmkomitees zum Thema relevanten Tagungen und Konferenzen, etc., ließen sich, sofern diese Daten in einem gemeinsam verwendeten Format wie dem Expertfinder Vokabular verfügbar sind, Kriterien für geeignete Experten in einer entsprechenden deklarativen Regel- oder Abfragesprache (wie RIF oder SPARQL) formulieren, oder bereits definierte Kriterien von früheren Events wiederverwenden. Unter Verwendung von akzeptierten Metadaten-Vokabularen für Wissenskategorien, Publikationen, etc., könnten Lehrmodell-Beispiele, wie in [19] präsentiert, die eine Kombination von OWL und Regeln zur Auswahl der Gutachter verwenden, tatsächlich im größeren Rahmen realisiert werden. Portale zum Management von Tagungen wie Easychair<sup>20</sup> könnten in Zukunft solche fortschrittlichen Features beinhalten um den Wissenschaftsprozess objektiver, durchschaubarer und einfacher verwaltbar zu machen.

---

<sup>19</sup> CORDIS bietet akademischen Institutionen, Klein- und Mittelbetrieben, bis hin zu weltweit operierenden Konzernen ein Forum, um Partner für gemeinsame Forschungs- und Entwicklungsprojekte innerhalb des siebten Rahmenprogramms der EU zu finden, <http://cordis.europa.eu/>.

<sup>20</sup> <http://www.easychair.org/>

### Datensicherheit und Vertrauensabkommen für personenbezogene Metadaten

In allen bisher genannten Szenarien ist es unabdingbar, Teile der veröffentlichten Metadaten vor unautorisiertem Zugriff zu schützen. Dies könnte beispielsweise durch den Austausch (zeitlich beschränkter) Zugangsschlüssel für verschlüsselte Metadaten oder mittels regelbasierter Verhandlung von Vertrauensabkommen (siehe z.B. [9, 10] oder das unter [21, Abschnitt 2.2] beschriebene Anwendungsszenario für weitere Details) realisiert werden. Regeln, wie die in Abschnitt 2, können so eine automatische Verhandlung zwischen Agenten steuern. Auf FOAF beschränkt, wäre es etwa denkbar, Mechanismen zu definieren, die sicherstellen, dass Personen, die meine Telefonnummer sehen wollen, ein Service aufrufen müssen, wo nur die Personen Zugang haben, welche ich kenne. Die Konfiguration solcher Services ließe sich mittels einer erweiterten FOAF Datei vornehmen. Diese Datei würde dann die eindeutigen URIs der Personen, die ich kenne (mittels `foaf:knows` Property), eine verschlüsselte Telefonnummer, sowie die Adresse eines Services zur Entschlüsselung enthalten. An dieser Stelle ist beispielsweise denkbar, dass das Service nur die bekannten Personen mittels Angabe ihrer OpenId<sup>21</sup> aufrufen könnten. Erweiterungen von FOAF um OpenId werden derzeit auf der FOAF Entwickler Mailingliste aktiv diskutiert<sup>22</sup>.

Die Erweiterung von FOAF um Verschlüsselung und Protokolle zum sicheren Metadaten austausch zwischen Agenten im Web ist nur ein Beispiel. Verschiedenste Erweiterungen von RDF Vokabularen um Digitale Signaturen, Zertifikate, oder beliebige komplexe Verhandlungsprotokolle wären an dieser Stelle denkbar.

Eine umfassende Beschreibung der ExpertFinder Anwendungsszenarien mit weiteren Links findet sich unter

<http://rdfweb.org/topic/ExpertFinderUseCases>.

Des Weiteren hat die ExpertFinder Initiative bisher zwei Workshops veranstaltet, deren Beiträge sich mit den Fragestellungen der semantischen Expertensuche im Web befassen (Details hierzu finden sich unter: <http://fews.semanticweb.org/>).

## 4 Das ExpertFinder Vokabular Framework

Wie bereits erwähnt, ist der Ansatz eine neue Ontologie zu entwickeln, eine aktive User Community zu akquirieren, bzw. eine neue Ontologie in einer bestehenden Community zu etablieren wenig erfolgversprechend. Statt eine vollkommen neue Ontologie zu entwickeln, die allen Anforderungen der Expertensuchdomäne genügt, schlagen wir daher ein Framework vor, das auf existierenden Vokabularen (Industriestandards, bestehender Klassifikationen und Taxonomien) basiert und diese

---

<sup>21</sup> <http://openid.net/>

<sup>22</sup> <http://lists.usefulinc.com/pipermail/foaf-dev/2007-September/008674.html>,  
<http://lists.usefulinc.com/pipermail/foaf-dev/2007-August/008648.html>,  
<http://lists.usefulinc.com/pipermail/foaf-dev/2007-June/008568.html>

sinnvoll miteinander kombiniert. Wie in Abbildung 1 dargestellt, lassen sich mehrere relevante “Komponenten” für die Beschreibung von Experten (d.h. Personen, Organisationen, Communities) identifizieren:

- *allgemeine Beschreibungsdaten* für Personen, Communities und Organisationen,
- *Beziehungen* zwischen Personen, Communities und Organisationen,
- vergangene und gegenwärtige *Aktivitäten und Projekte*,
- *Ausbildungsaspekte* (Schulbildung, Univeristätsabschlüsse, Kurse, etc.), sowie
- sonstige *Qualifikationen und Fähigkeiten*

Zusätzliche Bereiche wie *Ereignisse* (Konferenzen, Tagungen), *Publikationen, Meinungen und Bewertungen (Ratings)* durch Dritte sowie *Empfehlungen und Referenzen*, die nicht direkt mit der Beschreibung konkreter Personen oder Organisationen verbunden sind, aber aufgrund ihrer Relevanz für die Expertensuche abgedeckt sein sollten, werden ebenfalls in das Framework integriert. In all diesen Bereichen wollen wir unter Berücksichtigung der speziellen Anforderungen der Expertensuche die meist verbreiteten Vokabulare (i) identifizieren, (ii) überprüfen, inwieweit sie formalisiert sind, (iii) untersuchen, welche Überschneidungen und Überlappungen oder auch Widersprüche zwischen den bestehenden Ontologien bestehen, um letztendlich (iv) Richtlinien und Vorschläge zu erarbeiten, wie diese Vokabulare wiederverwendet, miteinander kombiniert und für die Beschreibung von und Suche nach Expertise optimal eingesetzt werden können.

Im nachfolgenden Abschnitt geben wir einen kurzen Überblick über die unser Meinung nach wichtigsten Ausgangsformate. Wir stellen dabei nicht den Anspruch, eine vollständige Liste aller unterschiedlichen Ontologien für die verschiedenen relevanten Bereiche zu erstellen, wollen aber weithin verbreitete und akzeptierte Formate abdecken.

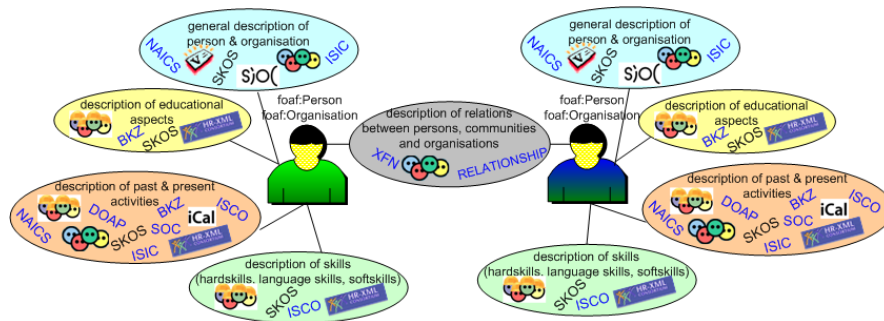


Abb. 1 Teilaspekte der Beschreibung von Expertise [2]

## 4.1 Ausgangspunkte: FOAF, SIOC & SKOS

Unser Überblick beginnt mit drei Projekten/Vokabularen, die als Ausgangspunkt unserer Arbeit dienen: Dies sind zum einen die Friend of a Friend (FOAF)<sup>23</sup> [13], sowie die Semantically-Interlinked Online Communities (SIOC)<sup>24</sup> [11, 6] Ontologien und zum anderen das Simple Knowledge Organisation System (SKOS)<sup>25</sup>. Diese drei Vokabulare dienen als solide Basis, die viele der erwähnten "Komponenten", die in der Expertenbeschreibung eine entscheidende Rolle spielen, abdeckt. Darüber hinaus vereinen sie Formate, die innerhalb von Online-Communities zunehmend Verbreitung finden und somit bereits auf eine große Anzahl von Anwendern verweisen können. Die Wiederverwendung dieser Vokabulare kommt unserem Vorhaben also sehr entgegen, erhöht die Wahrscheinlichkeit eine kritische Masse für das Framework zu erreichen, und erleichtert die Integration unseres Vokabulars in existierende Strukturen der Organisationen.

### Friend of a Friend (FOAF)

Die FOAF Ontologie [13], die ursprünglich als quasi "Metadatenbasis für eine maschinenlesbare Homepage" gedacht war, stellt Informationen/Metadaten über Menschen, Gruppen, Organisationen und andere relevante Konzepte bereit. FOAF erlaubt hauptsächlich Personen zu beschreiben zu (z.B.: was sie machen und wie sie miteinander interagieren). Eine der wichtigsten FOAF Properties ist, wie bereits in einem Beispiel erwähnt, das Attribut "foaf:knows". Durch das Einfügen von "foaf:knows"-Verknüpfungen kann man die Beziehungen zwischen Personen oberflächlich darstellen. Damit bietet FOAF einen einfachen Mechanismus um soziale Netzwerke abzubilden. Mit anderen Worten, die Aggregation von FOAF Daten aus vielen einzelnen persönlichen Homepages führt zur Bildung von verteilten sozialen Netzwerken, die wiederum mit größeren sozialen Onlinenetzwerken wie LiveJournal<sup>26</sup> oder Tribe<sup>27</sup> verbunden werden können. Wenn wir FOAF vom Blickwinkel der Expertenbeschreibung betrachten, finden wir eine Reihe relevanter Properties wie z.B.:

- foaf:interest definiert die Interessen einer Person.
- Mit foaf:publications oder foaf:Documents können Publikationen oder weitere Dokumente einer bestimmten Personen zugeordnet werden.
- Durch die Attribute foaf:topic und foaf:primaryTopic, die Publikationen/Dokumente einem bestimmten Thema zuweisen, lassen sich Interessen bestimmter Person ermitteln (Zuweisung eines Dokuments zu einer bestimmten Person, d.h. Autorenschaft, erfolgt über foaf:made/maker).

<sup>23</sup> <http://www.foaf-project.org/>

<sup>24</sup> <http://sioc-project.org/>

<sup>25</sup> <http://www.w3.org/2004/02/skos/>

<sup>26</sup> <http://rdfweb.org/topic/LiveJournal>

<sup>27</sup> <http://www.tribe.net>

- `foaf:currentProject/pastProject` gibt Aufschluss über gemeinsame oder individuelle Vorhaben und Projekte, an denen die bestimmte Person beteiligt ist oder war.<sup>28</sup>

Es existieren bereits mehrere Erweiterungen, Ergänzungen und zusätzliche Module für die FOAF Ontologie, die für die erwähnten Anwendungsszenarien der ExpertFinder Initiative relevant sind. An dieser Stelle sei beispielsweise ein Managementsystem für Benutzerprofile, FOAFRealm [28], erwähnt, das auf FOAF basiert und Funktionalitäten wie Authentisierung, Zugriffskontrolle, und Semantic Social Collaborative Filtering anbietet. Das System erlaubt Benutzern, die persönlichen Taxonomien mit anderen Nutzern zu teilen und diese entlang des sozialen Netzwerks unter Verwendung von WordNet<sup>29</sup>, DDC<sup>30</sup> und DMoz<sup>31</sup> als Basisklassifikationen zu annotieren. Darüber hinaus können User ihre Dokumente oder Bookmarks klassifizieren und anderen Benutzern erlauben, auf diese Ressourcen mit Hilfe von FOAFRealm zuzugreifen. Dies wird beispielsweise in einem System für den Dokumentenaustausch wie JeromeDL<sup>32</sup> – einer “semantischen” digitalen Bibliothek – implementiert. In FOAFRealm wird zudem jeder Benutzergruppierung ein Kompetenzwert zugewiesen, der die Qualität der Informationen reflektiert, die diese Gruppe bereitstellt. Der Kompetenzwert wird auf Basis von einem Algorithmus<sup>33</sup> berechnet, der das soziale Netzwerk analysiert. Dies verschafft dem Benutzer einen Überblick über die Qualität der Expertise anderer Beteiligter seines sozialen Netzwerks zum gegebenen Thema.

### Semantically-Interlinked Online Communities (SIOC)

Das SIOC Projekt [11, 6] stellt ein Framework zur Verknüpfung und zum Austausch von Informationen zur Verfügung, die aus Internet-basierten Diskussionsforen und Communityportalen kommen. Communities werden durch Benutzer gebildet, die themenbezogene Einträge verfassen, und an bestimmten Diskussionsforen teilnehmen, die von einer Vielzahl von Webseiten und Diskussionplattformen abonniert werden können – z.B. durch aktive Anmeldung oder über RSS (Really Simple Syndication Format). Die Basis der SIOC Ontologie bildet ein RDF-Schema, das die Hauptkonzepte einer Online Community beschreibt [11]. Auch wenn SIOC viel mehr Klassen und Properties beinhaltet, die generelle Notation ist wie folgt:

*Ein Benutzer (`sIOC:User`) erstellt einen Eintrag (`sIOC:Post`), der in einem Forum (`sIOC:Forum`) publiziert wird, das wiederum auf einer bestimmten Seite (`sIOC:Site`) zu finden ist.*

In Bezug auf die Expertensuche in sozialen Netzwerken ist `sIOC:topic` das wichtigste Property. Dieses Property definiert Ressourcen, die mit einem Eintrag in

<sup>28</sup> Die zeitliche Granularität dieser Zuordnung ist allerdings mit lediglich einer Unterscheidung zwischen gegenwärtigen und vergangenen Projekten sehr grob, mehr dazu später.

<sup>29</sup> <http://wordnet.princeton.edu/>

<sup>30</sup> <http://www.oclc.org/dewey/>

<sup>31</sup> <http://dmoz.de/>

<sup>32</sup> <http://www.jeromedl.org/>

<sup>33</sup> Dieser Algorithmus ähnelt dem Google zugrundeliegenden PageRank Algorithmus [14].

Verbindung gesetzt werden können. Durch Zusammenführung von allen Themen (`sIOC:topic`), die mit Einträgen eines bestimmten Benutzers über mehrere Webseiten assoziiert werden können, entsteht eine Art von Abbildung darüber, wo Interessengebiete und themenbezogene Expertise zu finden sind. Darüber hinaus können neben einzelnen Einträgen auch Foren (`sIOC:Forum`) oder Seiten mit bestimmten Themen assoziiert werden und, Benutzer, die Interesse an einem bestimmten Thema haben, können die Rolle des Abonnenten (`sIOC:subscriber_of`) für themenrelevante Diskussionskanäle annehmen.

### Simple Knowledge Organisation System (SKOS)

Das Simple Knowledge Organisation System (SKOS) [32] rundet die Basis unseres Frameworks ab. SKOS:

- erlaubt die Spezifikation von allgemeinen Konzepten und Bezeichnungen,
- definiert zusätzlich Eigenschaften von Bezeichnungen und Beziehungen zwischen Konzepten, die beschreiben, ob ein Konzept breiter oder enger (*broad/narrower*) gefasst ist als ein anderes,
- erlaubt bevorzugte oder alternative Bezeichnungen in verschiedenen Sprachen zu definieren,
- erleichtert sowohl die Repräsentation als auch die gemeinsame Nutzung von Terminologien, die nicht unbedingt die Ausdrucksstärke von Sprachen wie RDFS und OWL brauchen und, oder wo eine strenge durch `rdfs:subClassOf` definierbare Hierarchie nicht auferlegt werden kann.

In Bezug auf ExpertFinder und in Relation mit FOAF und SIOC kann SKOS (i) als Basis für die Definition und Zuordnung von Qualifikationen, (ii) für die Beschreibung von Expertisen- und Interessenbereichen (über `foaf:interest`) oder auch (iii) für die Beschreibung der Themen, die in SIOC-basierten online Communities diskutiert wurden (mittels `sIOC:topic`), benutzt werden. Außerdem ermöglicht der Einsatz von SKOS neben der Beschreibung von diskutieren Themen in Kombination mit zusätzlichen Regeln, die im nächsten Schritt in das ExpertFinder Framework eingebaut werden sollen, flexiblere Definition von Relationen zwischen verschiedenen SKOS-formalisierten Qualifikationen.

Um unnötige Duplikate oder Konflikte zwischen den Konzepten und Relationen in verschiedenen Vokabularen zu vermeiden, haben die Entwickler der SIOC Ontologie zusammen mit den Autoren von FOAF und SKOS am Konzeptenabgleich gearbeitet. Diese Kooperation hat sich unter anderem auf folgende Konventionen geeinigt:

- Das Konzept `sIOC:User` wurde als Subklasse von `foaf:onlineAccount` definiert, sodass die existierenden Properties aus FOAF wiederverwendet werden können, aber gleichzeitig auch neue Properties für Benutzer von Foren in SIOC definiert werden können, ohne direkte Auswirkungen auf die FOAF Ontologie zu haben.
- Wie in Abbildung 2 ersichtlich, erlaubt das Konzept `foaf:Person` mehrere `sIOC:User Profile` (über die Beziehung `foaf:holdsOnlineAccount`) mit einer Person zu assoziieren.

- Inhalte, die durch einen Benutzer (`sIOC:User`) in einem bestimmten Forum (e.g., Weblog, Mailingliste, Bulletin Board) veröffentlicht wurden, können über das Property `sIOC:topic` mit einem SKOS Konzept (`skos:Concept`) verlinkt werden (z.B.: vgl. Abb. 2 - ein Eintrag behandelt "Wolken" und ein weiterer bezieht sich auf ein engeres Konzept "Regenwolken").

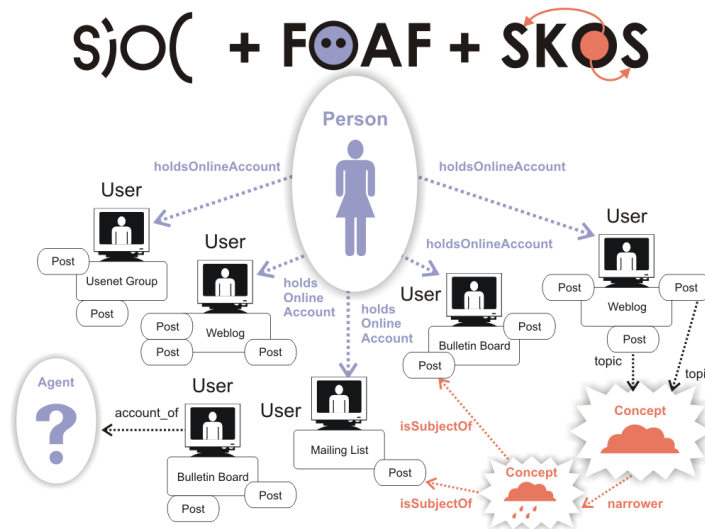


Abb. 2 Verbindungen zwischen SIOC, FOAF und SKOS [2]

## 4.2 Erweiterungen des ExpertFinder Frameworks

Zwar decken FOAF, SIOC und SKOS zum großen Teil die allgemeine Beschreibung von sowie Relationen zwischen Personen, Communities und Organisationen ab, doch fehlen nach wie vor entscheidende Komponenten, um das in Abb. 1 gezeigte Gesamtbild zu vervollständigen. An dieser Stelle wollen wir einige existierende Vokabulare und Standards, die die fehlenden Teile abdecken können, kurz erläutern. Was fehlt:

- In FOAF fehlen etwa detaillierte Informationen über Adressen, die durch komplementäre Standards wie *vCard* abgedeckt sind.
- Die einzige Beziehung zwischen Personen, die in FOAF definierbar ist, ist die `foaf:knows` Relation. Da wir zwischenmenschliche Beziehungen in unserem Framework genauer beschreiben wollen, werden wir die *RELATIONSHIP* & *XFN* Vokabulare integrieren.

- Projekte können mit Personen oder Gruppen durch die `foaf:currentProject` und `foaf:pastProject` Properties verlinkt werden. Allerdings, erscheint diese Beschreibung zu grobkörnig und zu allgemein, um Projekte genau annotieren zu können, wenn beispielsweise der exakte Zeitrahmen des Projektes von Bedeutung sein soll – Durch weiter zurückliegende Projekte erlangte Expertise ist oft weniger wichtig als unmittelbar abgeschlossene Projekte. Hier schlagen wir die Benutzung des *DOAP* (Description of a Project) Vokabulars vor.
- Detailliertere Informationen zum Lebenslauf können beispielsweise mit Hilfe der *DOAC* (Description of a Career) oder *BIO* Vokabulare spezifiziert werden.
- Im wissenschaftlichen Kontext sind oft die Publikationen der entscheidende Maßstab für Qualität und Niveau der Expertise einer bestimmten Person. Die Verbindung zwischen Personen und ihren Publikationen, die mit dem Konzept `foaf:Document` annotiert werden, kann durch die Properties `foaf:maker`, sein inverses Attribut `foaf:made` und `foaf:publications`<sup>34</sup> erstellt werden. Detaillierte Beschreibungen der jeweiligen Publikationen sind mit dem erwähnten Basis-Vokabular jedoch nicht abgedeckt. Diese Lücke kann durch *Bib-TeX*, ein de facto Standard im wissenschaftlichen Publikationsbereich, geschlossen werden.
- Informationen über Veranstaltungen und Veranstaltungsteilnehmer sind mit den existierenden Vokabularen nur unzureichend abgedeckt. Als de facto Standard wäre *iCal* ein naheliegender Kandidat für diese Aufgabe.
- SKOS bietet zwar ein allgemeines Framework für die Beschreibung von Qualifikationen und Interessen, dennoch fehlt eine konkrete Klassifikation, die verschiedene Themen und Interessensgebiete abdecken würde. Detaillierte Terminologien aus verschiedenen Domänen, die die Qualifikationen und Fähigkeiten einer Person beschreiben können, sind an dieser Stelle unabdingbar. Wir werden im folgenden einige mögliche Ausgangspunkte diskutieren.

### Verfeinerung von Personaldaten: vCard

*vCard* ist ein Standard zur Beschreibung von personenbezogenen Daten für beispielsweise Visitenkarten oder Adressbücher. Mittlerweile existieren verschiedene Repräsentationen von *vCard* Daten. Für unsere Zwecke ist allerdings hauptsächlich die RDF-basierte Syntax<sup>35</sup> relevant. Kontaktinformationen wie Telefonnummern oder E-Mail Adressen können mit Hilfe von *vCard* beispielsweise durch die Unterscheidung zwischen privaten und geschäftlichen Telefonnummern (`vCard:homeTel` und `vCard:workTel`) genauer und feingranularer ausgedrückt werden, als das mit dem FOAF Vokabular möglich wäre. An dieser Stelle sei angemerkt, dass sich Telefonnummern von *vCard* nicht einfach als Unterklassen von `foaf:phone` beschreiben lassen, weil *vCard* RDF Literalwerte wohingegen FOAF URIs benutzt, die mit dem `tel:` Schema qualifiziert werden. Um mit dieser Situation umgehen zu können, schlagen wir vor, entweder die FOAF Repräsentation in *vCard*/RDF zu adoptieren und die entsprechende Properties als Sub-

<sup>34</sup> `foaf:publications` verweist, im Gegensatz zu `foaf:made` auf eine Liste von Publikationen

<sup>35</sup> <http://www.w3.org/TR/vcard-rdf>, <http://www.w3.org/2006/vcard/ns>



properties von `foaf:phone` zu definieren, oder direkte Übersetzungsregeln für die entsprechende Konvertierung in einer geeigneten Regelsprache zu definieren, wie bereits im kleinen SPARQL Beispiel weiter oben vorgeschlagen. Solche Übersetzungsregeln müssen nicht immer bidirektional sein. Beispielsweise kann nicht jede `vCard:email` einfach auf eine `foaf:mbox` abgebildet werden, weil das `foaf:mbox` Property in FOAF entsprechend der Ontologiedefinition eindeutig einer bestimmten Person zugeordnet sein muss. Im Gegensatz dazu sind in vCard auch mehrdeutige Email Adressen zulässig. Ferner können Informationen über die organisatorischen Zugehörigkeit oder Rolle einer Person in einer Organisation auf eine bestimmte Expertise oder Grad des Wissens hindeuten, was wiederum mit SKOS Konzepten annotiert werden kann.

### **Verfeinerung von Relationen: RELATIONSHIP & XFN**

*RELATIONSHIP*<sup>36</sup> und *XFN (XHTML Friends Network)*<sup>37</sup> sind zwei Vokabulare, die für Beschreibung von zwischenmenschlichen Beziehungen benutzt werden können. Nachdem das FOAF Property `foaf:knows` die Relationen zwischen Personen für viele Anwendungen zu oberflächlich beschreibt, schließen diese Vokabulare diese Lücke durch Festlegung von detaillierteren Subproperties.

### **Verfeinerung von Projektbeschreibungen: Description of a Project (DOAP)**

*DOAP (Description of a Project)*<sup>38</sup> ist ein weiteres RDF Vokabular, das ursprünglich für die Beschreibung von Open Source Projekten gedacht war, das sich aber zunehmend auch für andere Projekte verbreitet hat. DOAP konzentriert sich dabei auf:

- die internationalisierte Beschreibung von Softwareprojekten und ihren Ressourcen;
- den Datenaustausch zwischen offenen Softwareverzeichnissen;
- automatische Konfiguration von Ressourcen wie CVS Repositories;
- Interoperabilität mit anderen bekannten Web Metadaten Projekten (RSS, FOAF, DC);
- die Erweiterungsmöglichkeit des Vokabulars für spezielle Zwecke.

DOAP beschreibt den aktuellen Zustand eines Projekt ohne jedoch explizit auf Änderungen und Updates anzugehen. Allerdings kann, um ein Repository auf dem neuesten Stand halten zu können, CodeZoo mit Atom<sup>39</sup> Feed samt eingebettetem DOAP Vokabular eingesetzt werden. Jedoch auch wenn so ein Feed für DOAP benutzt werden kann um ältere Versionen beizubehalten, ist eine Methode zur Transformation der Informationen nach RDF und zur Unterscheidung zwischen aktuellen und vergangenen Versionen dringend notwendig. DOAP verwendet das `foaf:Person` Konzept, um Personen, die am jeweiligen Teil des Projektes beteiligt sind, zu referenzieren. Trotz der relativ einfachen Integration der

<sup>36</sup> <http://vocab.org/relationship/>

<sup>37</sup> <http://gmpg.org/xfn/join>

<sup>38</sup> <http://usefulinc.com/doap/>

<sup>39</sup> [http://www.codezoo.com/about/doap\\_over\\_atom.csp](http://www.codezoo.com/about/doap_over_atom.csp)

beiden Vokabulare kann die genaue Dauer der Teilnahme einer Person an einem bestimmten Projekt nicht spezifiziert werden. Das Problem kann weder durch die Benutzung der FOAF Properties `foaf:project`, `foaf:currentProject` oder `foaf:pastProject` noch den Einsatz von DOAP gelöst werden. An dieser Stelle schlagen wir das Hinzufügen neuer Attribute für den Beginn und das Ende der Projektbeteiligung (möglicherweise unter Verwendung von iCal) vor. Auch die Benutzung von temporalem RDF [15] wäre an dieser Stelle denkbar.

### **Verfeinerung von typischen Lebenslaufdaten: DOAC, Resume RDF und BIO**

*DOAC (Description of a Career)*<sup>40</sup> ist ein RDF Vokabular für die Beschreibung von fachlichen Qualitäten und Erfahrungen wie man sie in Lebensläufen findet. Die Metadaten dieses Vokabulars erlauben die Beschreibung von und erleichtern Suche nach passenden Stellen bzw. Bewerbern im Sinne von Anforderungen eines Jobangebots. DOAC wurde als komplementäres RDF Vokabular zum Europäischen Lebenslauf (auch bekannt als Europass<sup>41</sup>) entwickelt, der aus einer FOAF+DOAC Datei generiert werden kann. DOAC beinhaltet Informationen über die Ausbildung, den beruflichen Werdegang, Publikationen, Sprachen und weitere Qualifikationen, die normalerweise in einem Lebenslauf Platz finden.

DOAC benutzt die FOAF Klasse `foaf:Person` für die allgemeine Beschreibung von Arbeitssuchenden und das Konzept `foaf:Organisation` für Schulen und andere Institutionen, die die jeweilige Person besucht hat und die in einem Lebenslauf eine Rolle spielen können. Darüber hinaus könnte das Konzept `foaf:pastProject` als Subklasse des Konzepts `doac:Experience` angesehen werden (wenn auch nicht explizit so im Vokabular definiert). Dies würde uns erlauben, nicht nur die Erfahrungen eines Jobsuchenden bei einer Firma als einzelnen Attributwert darzustellen, sondern fein-granulare Beschreibungen der Erfahrungen bezüglich verschiedener Projekte, an deren die Person während der Beschäftigungszeit bei der einer bestimmten Organisation beteiligt war, zu erstellen. Weiterhin mag man das Property `doac:publication`, das eine Beziehung zwischen einer Person (`foaf:Person`) und einer Publikation (`doac:Publication`) herstellt, als Subproperty von `foaf:made` bzw. als Verfeinerung von `foaf:publications` ansehen.

In ähnlicher Weise erweitert das *Resume RDF*<sup>42</sup> Schema FOAF Profile mit Lebenslaufinformationen und kann somit als Alternative zu DOAC gesehen werden. Das Schema beinhaltet Konzepte und Properties um berufliche und akademische Erfahrungen, Qualifikationen, Kurse und Zertifikate, Publikationen und weitere Referenzen annotieren zu können. Auch hier schlagen wir einen Link zum SKOS Vokabular vor, um für die Beschreibung entsprechender Konzepte bestehende Taxonomien wiederverwenden zu können.

---

<sup>40</sup> <http://ramonantonio.net/doac/>

<sup>41</sup> <http://europass.cedefop.europa.eu/>

<sup>42</sup> <http://purl.org/captsolo/semweb>

Das letzte Vokabular im Bereich von Lebenslaufinformationen ist das *BIO Vokabular*,<sup>43</sup> das minimale biographische Informationen über lebende und verstorbene Personen umfasst.

### **Verfeinerung von bibliographischen Informationen: BibTeX, DC, etc.**

*BibTeX*, das von Patashnik und Lampion 1985 als Format für die Verwaltung bibliographischer Daten in LaTeX entwickelt wurde [36], hat sich mittlerweile als de facto Standard für Veröffentlichungen bibliographischer Informationen besonders in online Publikationsdatenbanken (wie DBLP<sup>44</sup> und Citeseer<sup>45</sup>) etabliert. Da verschiedene RDF Versionen von BibTeX existieren (e.g. bibtex2rdf, bib2rdf, BibTeX/RDF<sup>46</sup>), die vorhandene Formate auf ähnliche Art und Weise wie ExpertFinder (wieder)verwenden, können diese direkt übernommen oder mit umfangreicheren und reichhaltigeren Ontologien für digitale Bibliotheken wie MarcOnt<sup>47</sup> kombiniert werden. Da MarcOnt selbst auch den Import und Export von BibTeX Daten ermöglicht, schlagen wir ein RDF Vokabular, das von bibtex2rdf (cf. [22]) unterstützt wird, vor. Zu guter Letzt sei die Dublin Core Metadaten Initiative (DCMI<sup>48</sup>) erwähnt, die vor 10 Jahren durch Bibliothekare mit dem Ziel der Erstellung von Metadaten für die Beschreibung von Dokumenten gestartet wurde. Das Vokabular von Dublin Core kann als Untermenge von BibTeX angesehen werden und wird beispielsweise in bibtex2rdf so weit wie möglich wiederverwendet.

### **Klassifikationen & Standards für Qualifikationen und Themengebiete**

Wie bereits erwähnt fehlen uns zur Komplettierung des ExpertFinder Frameworks nach wie vor konkrete Konzepthierarchien und Taxonomien für Themen und Kompetenzbereiche. Im folgenden Abschnitt beschreiben wir ausgewählte Standards und Klassifikationen von Berufen, Kompetenzen und ökonomischen Aktivitäten, die als solche Schemata für die Beschreibung von Qualifikationen und Themengebieten dienen könnten. Einige dieser Standards werden beispielsweise als Werkzeuge für die Präsentation von Statistiken über Ausbildung und Training auf nationalem als auch internationalem Level benutzt werden. Andere wiederum werden für die Förderung der internationalen Vergleichbarkeit von Daten bezüglich ökonomischer Phänomene und der internationalen Kommunikation eingesetzt.

- Das *Standard Occupational Classification (SOC)*<sup>49</sup> System wird von föderalen Statistikagenturen der USA benutzt, um Berufsgruppen in 820 Kategorien, 23 Hauptgruppen, 96 Untergruppen und 449 Tätigkeitsbereiche zu klassifizieren. Je-

<sup>43</sup> <http://vocab.org/bio/0.1/>

<sup>44</sup> <http://dblp.uni-trier.de/>

<sup>45</sup> <http://citeseer.ist.psu.edu/>

<sup>46</sup> <http://www.l3s.de/~siberski/bibtex2rdf/>,  
<http://www.cs.vu.nl/~mcaklein/bib2rdf/>,  
<http://zeitkunst.org/bibtex/0.1/>

<sup>47</sup> <http://www.marcont.org/>

<sup>48</sup> <http://dublincore.org/>

<sup>49</sup> <http://www.bls.gov/soc/>

der Tätigkeitsbereich beinhaltet detaillierte Berufe mit ähnlichen Qualifikations-, Ausbildungs- und Erfahrungsprofilen.

- Die Berufskennziffer (BKZ)<sup>50</sup> ist ein deutsches Pendant zum SOC System und klassifiziert Arbeitnehmer sehr feingranular in 5597 Berufskategorien.
- Das *International Standard Classification of Occupations (ISCO-88)* wurde entwickelt, um die internationale Kommunikation bezüglich Berufen und Berufsgruppen zu erleichtern. Personen werden hier nicht nur aufgrund einer einzelnen Tätigkeit sondern basierend auf ihre, in der Vergangenheit und Gegenwart ausgeübten klassifiziert.
- Die *International Standard Industrial Classification of All Economic Activities (ISIC)*<sup>51</sup> ist eine Standardklassifikation für ökonomische Aktivitäten, in der ebenfalls anhand der ausgeübten Aktivitäten klassifiziert wird.
- Das *North American Industry Classification System (NAICS)*<sup>52</sup> bietet eine allgemeine Definition der Industriezweige für Kanada, Mexiko und die USA und ist ursprünglich dafür konzipiert, ökonomische Analysen zu erleichtern.

Weitere Standards, die Produkte oder Dienstleistungen klassifizieren wie *eCI@ss*<sup>53</sup>, *eOTD*, oder *RosettaNet Technical Dictionary*<sup>54</sup> und *UNSPSC*<sup>55</sup> können ebenfalls in der Beschreibung von Qualifikationen und Themenbereichen eingesetzt werden. Allerdings sind fast alle der hier genannten Standards und Klassifikationen, mit wenigen Ausnahmen wie etwa *eCI@ssOWL* [24], derzeit leider (noch?) nicht in einer ontologisierten Form verfügbar. Abgesehen von speziellen Klassifikationssystemen, die noch web-fähig gemacht werden müssen, können online Enzyklopädien wie Wikipedia für die Themenklassifikation eingesetzt werden. Jüngste Versuche einer semantischen Strukturierung von Wikis (vgl. [30]) unterstützen genau diesen Ansatz. Für abgeschlossene, spezifische Domänen, wie etwa Informatik, können beispielsweise spezielle Kategorisierungen benutzt werden. Hier seien die ACM Kategorien<sup>56</sup>, oder die Semantic Web Themenhierarchie<sup>57</sup> als Beispiele genannt, die bereits URI-adressierbare Kategorien zur Verfügung stellen damit direkt als SKOS Ausdrücke einsetzbar sind. In diesem Sinne könnten wir das Beispiel vom Abschnitt 2 unter Verwendung einer spezifischeren Klassifikation wie folgt ändern/verfeinern:

*Alle aufgelisteten Personen unter der URI <http://www.w3.org/People/> sind Experten in [http://ontoworld.org/wiki/Category:Topic\\_Semantic\\_Web](http://ontoworld.org/wiki/Category:Topic_Semantic_Web)*

wobei wir folgendes hinzufügen können:

<sup>50</sup> [www.arbeitsamt.de/hst/markt/news/BKZ\\_alpha.txt](http://www.arbeitsamt.de/hst/markt/news/BKZ_alpha.txt)

<sup>51</sup> <http://unstats.un.org/unsd/cr/registry/isic-4.asp>

<sup>52</sup> <http://www.census.gov/epcd/www/naics.html>

<sup>53</sup> <http://www.eclass.de/>

<sup>54</sup> <http://portal.rosettanet.org/cms/sites/RosettaNet/Standards/RStandards/dictionary/technical/index.html>

<sup>55</sup> <http://www.unspsc.org/>

<sup>56</sup> <http://www.acm.org/class/1998/>

<sup>57</sup> [http://ontoworld.org/wiki/Semantic\\_Web\\_Topic\\_Hierarchy](http://ontoworld.org/wiki/Semantic_Web_Topic_Hierarchy)

*[http://ontoworld.org/wiki/Category:Topic\\_Semantic\\_Web](http://ontoworld.org/wiki/Category:Topic_Semantic_Web)  
is skos:narrower than  
[http://en.wikipedia.org/wiki/Semantic\\_Web](http://en.wikipedia.org/wiki/Semantic_Web).*

### **Veranstaltungsdaten und temporale Informationen**

iCal [16] als de facto Standard für Kalender- und Eventinformationen, der von vielen Anwendungen und Applikationen benutzt und unterstützt wird, ist ein naheliegender Kandidat für ExpertFinder, wenn es darum geht, Verweise auf Events, Veranstaltungen oder zeitliche Intervalle zu beschreiben. Für den Einsatz von iCal spricht auch die Tatsache, dass bereits RDF Formate und Konvertierung für iCal verfügbar sind<sup>58</sup>. Allerdings ist anzunehmen, dass der Wortschatz von iCal alleine nicht ausreichend sein wird (beispielsweise für die Annotation der Gültigkeitsdauer bestimmter RDF Informationen wie die Dauer der Teilnahme an einem Projekt). Die Benutzung von RDF Erweiterungen um temporale Informationen [15] wäre eine alternative interessante Möglichkeit um die Gültigkeitsdauer von Tripeln oder Graphen auszudrücken. Diese Ideen nehmen allerdings noch keinen festen Platz im Portfolio etablierter Standards ein.

### **Beschreibung von Überlappungen und Mappings**

Die einzelnen vorgestellten Vokabulare bilden ein loses, teilweise überlappendes Framework von Ontologien und RDF Vokabularen und Kategorisierungen, um die in Abbildung 1 genannten Aspekte abdecken zu können. Der Schwerpunkt lag hier darauf, bestehende Formate weitestgehend wiederzuverwenden und einen Überblick zu geben, inwieweit diese kombiniert werden können. Dennoch decken die genannten Formate viele Aspekte doppelt ab, bzw. es gibt mehrerer Möglichkeiten unter Verwendung verschiedener der genannten Formate ein und den selben Sachverhalt auszudrücken. Die ExpertFinder Initiative zielt darauf ab, für jeden dieser mehrdeutigen Aspekte ein bestimmtes designiertes Format vorzuschlagen und Mappings von und nach evtl. überlappenden anderen Formaten zu definieren. Eine vorläufige, unvollständige Liste von Mappings und Überschneidungen der Konzepte und Attribute in den vorgestellten Vokabularen ist in Arbeit und unter [http://www.rdfweb.org/topic/ExpertFinder\\_2fmappings](http://www.rdfweb.org/topic/ExpertFinder_2fmappings) zu finden. Die Autoren sind über etwaige Anregungen zur Vervollständigung dieser Liste dankbar. Das Nichtvorhandensein eines dezidierten Standards zur Beschreibung solcher Mappings wurde bereits durch Vorschläge entsprechender Erweiterungen für SPARQL [39] in Angriff genommen.

---

<sup>58</sup> <http://www.w3.org/2002/12/cal/>,  
<http://www.kanzaki.com/courier/ical2rdf>,  
<http://torrez.us/ics2rdf/>

## 5 Relevante Projekte

Mehrere Projekte im Semantic Web Bereich haben bereits ihre eigenen Ontologien entwickelt um Personen, Organisation und Aktivitäten zu beschreiben. Beispielsweise die *KnowledgeWeb Plattform Ontologien*<sup>59</sup> [20], die *AKT Portal Ontologie*<sup>60</sup>, die *SWRC Portal Ontologie*<sup>61</sup> [42], oder die Ontologie der *DERI Semantic Web Portal (SWP) Arbeitsgruppe*<sup>62</sup> [34] decken viele Aspekte der Expertensuche ab und könnten somit als adäquate Ausgangspunkte angesehen werden. Allerdings scheinen diese Ontologien sich nur mäßig außerhalb der Projekte, in denen sie entwickelt wurden, verbreitet zu haben und bei ihrer Erstellung wurden existierende Quellen und de facto Standards zu einem weit geringeren Grad wiederverwendet, als wir es hier vorgeschlagen haben. Die DERI SW-Portal Ontologie basiert als einzige aus dieser Liste zum Teil auf FOAF, RSS und BibTeX.

Im Gegensatz zu den oben genannten Ontologien, die in den jeweiligen Projekten von Grund auf neu entwickelt wurden, setzen andere Projekte wie Wissensnetze, SemDis, FindXpRT und SIOC zu einem höheren Grad auf die Wiederverwendung existierender Vokabulare. Diese Projekte, auf die wir im folgenden etwas näher eingehen wollen, behandeln auch sehr ähnliche Anwendungsszenarien, wie wir sie in Abschnitt 3 beschrieben haben.

### Wissensnetze

Das Projekt *Wissensnetze*<sup>63</sup> analysiert anhand von konkreten Anwendungsszenarien das Potential semantischer Technologien zur Veränderung elektronischer Märkte und branchenspezifischer Wertschöpfungsketten. Grundlage der Szenarioentwicklung ist die Kombination verschiedener Anwendungsbereiche im Lichte der neuen Technologien. Die auf diese Weise entwickelten "Zukunftsszenarien" sind anschließend Gegenstand von Experimenten und Analysen. Das erste Szenario, das im Rahmen des Projektes untersucht und prototypisch entwickelt wurde, kommt aus der e-Recruitment Domäne und befasst sich mit online Jobsuche in einem "semantischen" Jobportal [5, 33, 43]. Im Human Resource Szenario (HR-Szenario) bilden Ontologien eine wesentliche Grundlage für die Beschreibung von Jobangeboten und -anfragen und deren semantischem Vergleich. Anhand verschiedener Kostenfaktoren, nach einem Kostenmodell für Ontologieentwicklung (ONTOCOM) [41], wurde entschieden, dass die Wiederverwendung von existierenden Ontologien kostengünstiger als eine Neuentwicklung der HR-Ontologie ist. Insgesamt wurden 24 Klassifikationen und Standards aus der HR-Domäne untersucht, von denen schließlich sechs angepasst und in der HR-Ontologie wiederverwendet wurden (SOC<sup>64</sup>, BKZ<sup>65</sup>,

<sup>59</sup> <http://knowledgeweb.semanticweb.org/semanticportal/OWL/>

<sup>60</sup> <http://www.aktors.org/ontology/portal>

<sup>61</sup> <http://swrc.ontoware.org/ontology>

<sup>62</sup> <http://sw-portal.deri.org/ontologies/swportal>

<sup>63</sup> <http://wissensnetze.ag-nbi.de>

<sup>64</sup> <http://www.bls.gov/soc/>

<sup>65</sup> [http://infobub.arbeitsagentur.de/download/public/dkz\\_daten/](http://infobub.arbeitsagentur.de/download/public/dkz_daten/)

WZ2003<sup>66</sup>, NAICS<sup>67</sup>, HR-XML<sup>68</sup> und die Skill Ontology[1]) um die Job- und Stellenangebote annotieren und diese einfacher in die industrielle Umgebung integrieren zu können. Wiederverwendung von Standards und Klassifikationen trägt hier zur Realisierung von leistungsfähigeren und flexibleren Lösungen im e-Recruitment Szenario bei.

### **SemDis**

Das *SemDis* Projekt adressiert die Entwicklung von Techniken für die Abfrage/Ermittlung von semantischen Relationen zwischen Personen. Beispielsweise werden `dblp:co-authorship` und `foaf:knows` benutzt um mögliche Interessenkonflikte zwischen Gutachtern und Autoren in wissenschaftlichen Begutachtungsprozess zu erkennen [4]. Eine Erweiterung dieser Arbeit zielt auf Ermittlung von möglichen Gutachtern durch Vergleich ihrer Expertise mit Themen einer Publikation, die sie beurteilen sollen. Die Expertise von Personen bezüglich verschiedener Themen oder Bereiche kann mit Hilfe der SwetoDblp Ontologie<sup>69</sup> beschrieben werden. SwetoDblp ist eine Ontologie basierend auf der bereits erwähnten DBLP Online Publikationsdatenbank.

### **FindXpRT**

Das Projekt *FindXpRT* (Find an eXpert via Rules and Taxonomies) [29] konzentriert sich auf Regelaspekte im Sinne der Kombinationen von FOAF Daten mit Regeln, die in RuleML [8] spezifiziert sind. Das implementierte System<sup>70</sup> erlaubt Benutzern FOAF Daten durch Anwendung von personenbezogenen Regeln entweder vor der Publikation (als eine Art von Filter) oder, auf Wunsch, aufgrund von bereits publizierten RuleML+FOAF Seiten (als Sichten/Views) abzuleiten.

### **SEEMP**

Das SEEMP (Single European Employment Market Place)<sup>71</sup> Projekt hat zum Ziel, eine Infrastruktur bereitzustellen, um Interoperabilität zwischen öffentlichen und privaten Arbeitsmarktdiensten über europäische Grenzen hinweg zu ermöglichen [45]. Der Ansatz von SEEMP basiert auf zwei Hauptkonzepten: Dienste und Semantik. Die Dienste, die von den Marktteilnehmern angeboten wurden, werden als spezielle Web Services modelliert. Die Semantik von Inhalten, die durch das System verwaltet wird, wurde in Ontologien verpackt, die wiederum als Basis für den Datenaustausch benutzt werden. Auch bei der Entwicklung der SEEMP Referenz Ontologie wurde soweit wie möglich auf die Wiederverwendung von existierenden Standards (ökonomische Aktivitäten: NACE Rev. 1.1<sup>72</sup>, Berufsbezeichnungen: ISCO-88

<sup>66</sup> <https://www-ec.destatis.de/csp/shop/sfg/bpm.html.cms.cBroker.cls?cmspath=struktur,vollanzeige.csp&ID=1018484>

<sup>67</sup> <http://www.census.gov/epcd/www/naics.html>

<sup>68</sup> Entwickelt vom HR-XML Consortium, <http://www.hr-xml.org>

<sup>69</sup> <http://lsdis.cs.uga.edu/projects/semdis/swetodblp/>

<sup>70</sup> <http://www.ruleml.org/usecases/foaf/JieLiMCSThesis.pdf>

<sup>71</sup> <http://www.seemp.org>

<sup>72</sup> [http://www.fifoost.org/database/nace/nace-de\\_2002c.php](http://www.fifoost.org/database/nace/nace-de_2002c.php)

(COM)<sup>73</sup>, ONET<sup>74</sup> und ISCED97<sup>75</sup>, Qualifikationen: EURES Qualifikationklassifikation<sup>76</sup>, usw.) gesetzt.

## 6 Zusammenfassung und Ausblick

Im Verlauf dieses Kapitels wurden Rahmenbedingungen zur Kombination, Wiederverwendung und Erweiterung bestehender RDF Vokabulare im Social Semantic Web diskutiert und die Vision des ExpertFinder Vokabular Framework vorgestellt. Wir haben existierende Vokabulare und Formate analysiert, die der Spezifikation von Personen, Communities und Organisationen samt deren Expertise dienen können. Außerdem haben wir uns mit der Problematik der Kombination und Erweiterungen verschiedener Quellformate auseinander gesetzt, wobei wir die Einschränkungen von FOAF, SIOC und SKOS, die wir als Ausgangspunkte betrachten, durch die Kopplung mit anderen existierenden Vokabularen zumindest teilweise auflösen konnten. Darüber hinaus, haben wir die Möglichkeit der Benutzung von Regelsprachen zur Beschreibung von Formatüberlappungen, sowie deren praktischer Verwendbarkeit zur Erstellung eines Ontologie-Netzwerks zur Beschreibung von Experten kurz motiviert und anhand von Beispielen vorgestellt. Desweiteren stellten wir einige Projekte genauer vor, die neue Ontologien in den für uns relevanten Bereichen auf Basis von Wiederverwendung vorhandener Vokabulare entwickelt haben. Alle Untersuchungen bezüglich des ExpertFinder Frameworks, die in dieser Arbeit präsentiert wurden, wurden im Hinblick auf die kritischen Erfolgskriterien (gemeinsam verwendete, maschinenlesbare Formate, kritische Masse von Anwendern/Nutzern und zusätzlich benötigte Technologien) durchgeführt, die wir eingangs definiert haben. Obwohl die Arbeit am ExpertFinder Framework sich gerade erst in der Anfangsphase befindet, können wir mit dem vorgestellten Vokabular bereits wesentliche Teile der vorgestellten Anwendungsszenarien abdecken. Zu guter Letzt sei gesagt, dass das vorgestellte Vokabular in Kombination mit Regeln für Formatüberlappungen nicht auf die erwähnten Szenarien beschränkt bleiben muss. Das ExpertFinder Vokabular Framework kann in unterschiedlichen Domänen rund um die Beschreibung und Suche von Expertisen und Personen benutzt werden, auch wenn wir noch intensiv an Mechanismen arbeiten müssen, die Datensicherheit und Vertrauensgarantien von Inhalten sicherstellen. Diese Aspekte scheinen derzeit in der Semantic Web Architektur generell noch nicht ausreichend abgedeckt.

**Acknowledgements:** Die Autoren danken den Mitgliedern der ExpertFinder Initiative für ihren Beitrag zum vorliegenden Artikel. Speziell danken wir Boanerges Aleman-Meza, Uldis Bojars, Harold Boley, John G. Breslin, Lyndon J.B. Nixon und Anna V. Zhdanova, die als Mitverfasser an einem Tagungsbeitrag [2] mitgewirkt haben, auf dem dieses Kapitel basiert.

---

<sup>73</sup> <http://www.gesis.org/dauerbeobachtung/gml/Daten/mz/allgemein/isco88com.pdf>

<sup>74</sup> <http://std.en.vicdir.com/1021>

<sup>75</sup> [http://www.uis.unesco.org/ev.php?ID=3813\\_201&ID2=DO\\_TOPIC](http://www.uis.unesco.org/ev.php?ID=3813_201&ID2=DO_TOPIC)

<sup>76</sup> <http://europa.eu.int/eures/home.jsp?lang=de>



Die Arbeit von Axel Polleres wurde durch die Europäische Kommission im Rahmen des *inContext* Projekts (IST-034718) sowie durch die Science Foundation Ireland im Rahmen des *DERI-Lion* Projekts (SFI/02/CE1/1131) gefördert. Die Arbeit von Malgorzata Mochol wurde durch das Projekt *Wissensnetze*, ein Teilprojekt im Berliner Forschungszentrum InterVal (Internet and Value Chains) im Rahmen des Projektes Internetökonomie vom Bundesministerium für Bildung und Forschung (BMBF), gefördert.

## Literaturverzeichnis

1. Y. Alan. Konstruktion der KOWIENOntologie. Projektbericht 2, Institut für Produktion und Industrielles Informationsmanagement, Universität Duisburg-Essen, 2003.
2. B. Aleman-Meza, U. Bojars, H. Boley, J. G. Breslin, M. Mochol, L. J.B. Nixon, A. Polles, and A. V. Zhdanova. Combining RDF Vocabularies for Expert Finding. In *4th European Semantic Web Conference (ESWC2007)*, pages 235–250, Innsbruck, Austria, June 2007. Springer.
3. F. Alkhateeb, J. Baget, and J. Euzenat. Extending SPARQL with Regular Expression Patterns. Technical Report 6191, Institut National de Recherche en Informatique et Automatique (INRIA), May 2007.
4. B. Aleman-Meza, et al. Semantic Analytics on Social Networks: Experiences in Addressing the Problem of Conflict of Interest Detection. In *15th International World Wide Web Conference (WWW2006)*, 2006.
5. C. Bizer, R. Heese, M. Mochol, R. Oldakowski, R. Tolksdorf, and R. Eckstein. The Impact of Semantic Web Technologies on Job Recruitment Processes. In *7. Internationale Tagung Wirtschaftsinformatik 2005*, pages 1367–1383, 2005.
6. Uldis Bojars, John G. Breslin, Diego Berrueta, Dan Brickley, Stefan Decker, Sergio Fernández, Christoph Görn, Andreas Harth, Tom Heath, Kingsley Idehen, Kjetil Kjernsmo, Alistair Miles, Alexandre Passant, Axel Polleres, Luis Polo, and Michael Sintek. SIOC Core Ontology Specification, June 2007. W3C member submission, <http://www.w3.org/Submission/sioc-spec/>.
7. H. Boley, M. Kifer, P.-L. Pătrânjan, and A. Polleres. Rule Interchange on the Web. In *Reasoning Web 2007*, pages 269–309. Springer, September 2007.
8. H. Boley, S. Tabet, and G. Wagner. Design Rationale of RuleML: A Markup Language for Semantic Web Rules. In *Semantic Web Working Symposium (SWWS'01)*, pages 381–401. Stanford University, July/August 2001.
9. P. A. Bonatti and D. Olmedilla. Semantic Web Policies: Where are we and What is still missing? Tutorial at the European Semantic Web Conference (ESWC), June 2006.
10. P. A. Bonatti and D. Olmedilla. Rule-Based Policy Representation and Reasoning for the Semantic Web. In *Reasoning Web 2007*, pages 240–268. Springer, September 2007.
11. J. G. Breslin, A. Harth, U. Bojars, and S. Decker. Towards Semantically-Interlinked Online Communities. In *2nd European Semantic Web Conference (ESWC '05)*, pages 500–514, 2005.
12. D. Brickley and R. V. Guha (eds.). RDF Vocabulary Description Language 1.0: RDF Schema, February 2004. W3C Recommendation, 10 February 2004. Available from <http://www.w3.org/TR/rdf-schema/>.
13. D. Brickley and L. Miller. FOAF Vocabulary Specification, July 2005. <http://xmlns.com/foaf/0.1/>.
14. S. Brin and L. Page. The anatomy of a large-scale hypertextual Web search engine. In *7th International World Wide Web Conference (WWW1998)*, pages 107–117, 1998.
15. A. Vaisman C. Gutierrez, C. Hurtado. Temporal RDF. In *2nd European Semantic Web Conference (ESWC'05)*, 2005.
16. F. Dawson and D. Stenerson. Internet Calendaring and Scheduling Core Object Specification (iCalendar). <http://www.ietf.org/rfc/rfc2445.txt>, 1998.

17. D. Connolly (ed.). Gleaning Resource Descriptions from Dialects of Languages (GRDDL), July 2007. W3C Proposed Recommendation, <http://www.w3.org/TR/grddl/>.
18. J. Clark (ed.). XSL Transformations (XSLT) Version 1.0, November 1999. W3C Recommendation, <http://www.w3.org/TR/xquery/>.
19. T. Eiter, G. Ianni, A. Polleres, and R. Schindlauer. Answer Set Programming for the Semantic Web. Tutorial at the European Semantic Web Conference (ESWC), June 2006. Tutorial at the European Semantic Web Conference (ESWC), slides available at <http://asptut.gibbi.com/>.
20. E. Franconi. The Knowledge Web Network of Excellence. [http://interop-esa05.unige.ch/INTEROP/Proceedings/IST/IST1\\_kweb-.pdf](http://interop-esa05.unige.ch/INTEROP/Proceedings/IST/IST1_kweb-.pdf), 2005.
21. A. Ginsberg, D. Hirtle, F. McCabe, and P.-L. Patranjan (eds.). RIF Use Cases and Requirements. W3C Working Draft.
22. G. A. Grimnes, S. Schwarz, and L. Sauer mann. RDFHomepage or “Finally, a use for your FOAF file”. In *2nd Workshop on Scripting for the Semantic Web (SFSW '06)*, June 2006.
23. Andreas Harth, Jürgen Umbrich, and Stefan Decker. MultiCrawler: A Pipelined Architecture for Crawling and Indexing Semantic Web Data. In *5th Int.l Semantic Web Conference*, Athens, GA, USA, November 2006.
24. M. Hepp. Products and Services Ontologies: A Methodology for Deriving OWL Ontologies from Industrial Categorization Standards. *International Journal on Semantic Web & Information Systems*, 1(2):72–99, 2006.
25. D. Huynh, S. Mazzocchi, and D. Karger. Piggy Bank: Experience the Semantic Web Inside Your Web Browser. In *International Semantic Web Conference 2005 (ISWC2005)*, 2005.
26. R. Iannella. Representing vCard Objects in RDF/XML, February 2001. W3C Note, available at <http://www.w3.org/TR/vcard-rdf>.
27. G. Klyne and J. J. Carroll (eds.). Resource Description Framework (RDF): Concepts and Abstract Syntax, February 2004. W3C Recommendation, <http://www.w3.org/TR/rdf-concepts/>.
28. S. R. Kruk and S. Decker. Semantic Social Collaborative Filtering with FOAFRealm. In *Semantic Desktop Workshop co-located with ISWC2005*, 2005.
29. J. Li, H. Boley, V. C. Bhavsar, and J. Mei. Expert Finding for eCollaboration Using FOAF with RuleML Rules. <http://www.ruleml.org/papers/FindXpRT.pdf>.
30. S. Schaffert M. Krötzsch and D. Vrandečić. Reasoning in Semantic Wikis. In *Reasoning Web 2007*, pages 310–329. Springer, September 2007.
31. A. Malhotra, J. Melton, and N. Walsh (eds.). XQuery 1.0 and XPath 2.0 Functions and Operators, January 2007. W3C Recommendation, <http://www.w3.org/TR/xpath-functions/>.
32. A. Miles and D. Brickley (eds.). SKOS Core Vocabulary Specification, November 2005. W3C Working Draft, <http://www.w3.org/TR/swbp-skos-core-spec>.
33. M. Mochol, R. Oldakowski, and R. Heese. Ontology-based Recruitment Process. In *GI Conference (GI2004)*, 2004.
34. K. Möller, L. Predoiu, and D. Bachlechner. Portal Ontology - Semantic Web Portal Project. Technical Report Project Deliverable, V1.2., DERI Research Report, 2004.
35. C. Morbidoni, A. Polleres, and G. Tummarello. Who the FOAF knows Alice? A needed step towards Semantic Web Pipes. In *ISWC 2007 Workshop on New forms of Reasoning for the Semantic Web: Scaleable, Tolerant and Dynamic*, Busan, Korea, November 2007.
36. O. Patashnik. BIBTeXing, 1998.
37. P. F. Patel-Schneider, P. Hayes, and I. Horrocks (eds.). OWL Web Ontology Language Semantics and Abstract Syntax, February 2004. W3C Recommendation, <http://www.w3.org/TR/owl-semantics/>.
38. A. Polleres and C. Feier and A. Harth. Rules with Contextually Scoped Negation. In *3rd European Semantic Web Conference (ESWC2006)*, Budva, Montenegro, June 2006. Springer.
39. A. Polleres, F. Scharffe, and R. Schindlauer. SPARQL++ for Mapping between RDF Vocabularies. In *6th International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE 2007)*, Vilamoura, Algarve, Portugal, November 2007.

40. E. Prud'hommeaux and A. Seaborne (eds.). SPARQL Query Language for RDF, June 2007. W3C Candidate Recommendation, <http://www.w3.org/TR/2007/CR-rdf-sparql-query-20070614/>.
41. E. Paslaru Simperl, C. Tempich, and M. Mochol. *Technologies for Business Information Systems*, chapter Cost estimation for ontology development: applying the ONTOCOM model, pages 432–447. Springer, 2007.
42. Y. Sure, S. Bloehdorn, P. Haase, J. Hartmann, and D. Oberle. The SWRC Ontology - Semantic Web for Research Communities. In *12th Portuguese Conference on Artificial Intelligence - Progress in Artificial Intelligence (EPIA 2005)*, volume 3803, pages 218–231, 2005.
43. R. Tolksdorf, M. Mochol, R. Heese, R. Eckstein, R. Oldakowski, and C. Bizer. Semantic-Web-Technologien im Arbeitsvermittlungsprozess. *Wirtschaftsinformatik: Internetoekonomie*, 48(1):17–26, 2006.
44. Giovanni Tummarello, Renaud Delbru, and Eyal Oren. Sindice.com: Weaving the Open Linked Data. In *6th Int.l Semantic Web Conf. (ISWC2007)*, Busan, Korea, November 2007.
45. E. Della Valle and et. al. SEEMP: A Semantic Interoperability Infrastructure for e-Government Services in the Employment Sector. In *European Semantic Web Conference '07 (ESWC2007)*, pages 220–234. Springer, 2007.
46. A versit Consortium. vCard: The Electronic Business Card, 1997. <http://www.imc.org/pdi/vcardwhite.html>.