

Recrutare web 2.0

Vlad Posea

Universitatea Politehnica Bucuresti

Splaiul Independenței 313, sector 6

vlad.posea@cs.pub.ro

REZUMAT

Articolul descrie primii pași în realizarea unui sistem semantic de management al competențelor. Sistemul este destinat îmbunătățirii activității unei companii de recrutare specializate pe domeniul IT și utilizează tehnologii din domeniul webului semantic și al serviciilor web.

Cuvinte cheie

Semantic web, competente, ontologii, linkedin

Clasificare ACM

H.3.4 Systems and Software.

INTRODUCERE

Problema managementului și modelării competențelor este una destul de veche. Utilitatea obținerii unui astfel de model ar fi extrem de mare pentru realizarea unor instrumente software pentru gestiunea personalului sau pentru recrutare. Un astfel de instrument ar putea ajuta la planificarea training-ului în companie oferind o imagine a competențelor ce pot fi îmbunătățite sau ar putea ajuta la realizarea unei potriviri mai bune între cv-ul unui candidat și descrierea unui job la care acel candidat aplică.

Lucrarea de față descrie primii pași în realizarea unui sistem bazat pe un model al competențelor, sistem ce va fi folosit într-o companie de recrutare.

Un alt aspect ce va fi discutat în această lucrare se referă la utilizarea LinkedIn.com. LinkedIn este un site pe care oamenii își pot publica CV-urile și pe care se pot conecta cu colegii sau cu cunoștințe profesionale. LinkedIn este un instrument ce poate fi folosit în procesul de recrutare întrucât este un sistem prin care oamenii pot să-și mențină ușor la zi un CV într-un format oarecum standardizat. Un capitol din lucrare va descrie datele puse la dispoziție de LinkedIn.

DESPRE MODELAREA ȘI MANAGEMENTUL COMPETENȚELOR

Mai mulți autori au dorit realizarea unei ontologii a competențelor [1,2,3,4]. Printre ideile cheie menționate în aceste lucrări poate cea mai importantă ar fi cea referitoare la sursele de date. Astfel, o ontologie a competențelor ar trebui să fie bazată pe standarde din domeniul resurselor umane (HR), pe standarde ocupaționale și pe o descriere a domeniului în care ne propunem să modelăm competențele.

Surse de date menționate ar fi HR-XML¹, SOC (Standard occupational classification) – foarte util mai ales pentru domeniul IT², sau alte ontologii ca de exemplu KOWIEN [1]– ontologii care sunt menționate în literatură dar nu sunt disponibile pentru descărcare sau utilizare pe web.

HR-XML este un consorțiu ce se ocupă cu realizarea unor formate xml standard pentru utilizare în domeniul resurselor umane.

Având la dispoziție aceste surse de date încercăm să analizăm și definițiile competenței conform autorilor citați. Aceste definiții par să convergă către următoarea. O competență este un ansamblu de cunoștințe teoretice (knowledge), de cunoștințe practice (know-how) și de elemente de comportament (soft-skills) ce pot fi folosite într-un context specific pentru a atinge un obiectiv.

Din această definiție se desprinde faptul că pentru a modela competențele dintr-un domeniu este necesar să modelăm domeniul, să modelăm activitățile din acel domeniu și să utilizăm un model global al soft-skills.

Pe lângă definiție, lucrările citate oferă și un model al competențelor, model care poate fi rezumat prin următoarele idei:

- Modele complexe
- Competența conține cunoștințe, know-how/skill, comportament și câteodată context
- Modelul depinde de ceea ce se dorește ca aplicația să realizeze. Un model pentru o aplicație de recrutare este diferit de un model pentru o aplicație de gestiune a personalului
- Relațiile între competențe apar în model
- Sunt foarte importante relațiile între competențe și domeniul de lucru.

UTILIZAREA LINKEDIN CA SURSA DE DATE PENTRU DOMENIUL HR

LinkedIn este un site lansat în mai 2003 și care în acest moment este cea mai mare rețea profesională de pe glob, având peste 161 de milioane de membri în 200 de țări, din care 34 de milioane în Europa³ și 600.000 în România în ianuarie 2012⁴.

¹ <http://www.hr-xml.org>

² <http://www.bls.gov/soc/2010/soc151132.htm>

³ <http://press.linkedin.com/about>

⁴ <http://www.slideshare.net/amover/linkedin-demographics-statistics-jan-2012>

LinkedIn oferă o parte din informațiile de pe site prin intermediul unor servicii web sub formă de fișiere xml sau json. Pentru a accesa aceste servicii trebuie ca un utilizator să ofere unei aplicații un token care permite aplicației să facă cereri în numele utilizatorului. Cu acest token aplicația poate cere informații despre proprietarul tokenului, despre conexiunile acestuia (nu și despre conexiunile de nivel 2 sau mai mare) și poate efectua căutări ce vor returna informații publice despre alți utilizatori. Pentru a obține informații complete despre cât mai mulți utilizatori avem nevoie ca un număr cât mai mare de oameni să ne permită să efectuăm căutări în numele lor.

Datele pe care linkedin le oferă pot fi rezumate prin următorul URL care pune la dispoziție datele din profilul personal al utilizatorului:

```
http://api.linkedin.com/v1/people/~:(id,first-name,last-name,industry,headline,location:(name),num-connections,num-connections-capped,specialties,interests,date-of-birth,skills:(id,skill,proficiency,years),public-profile-url,positions,educations,certifications,relation-to-viewer:(distance),picture-url,main-address,languages:(id,language,proficiency))
```

URL-ul de mai sus preia informațiile personale de bază referitoare la utilizator (id, nume, prenume, funcție, loc unde se află, date despre numărul de conexiuni, despre domeniile de interes, despre cunoștințe (skills), experiență, educație, certificări.

Principalele probleme ale datelor oferite de LinkedIn sunt legate de faptul că utilizatorii au o mare libertate de completare a profilului propriu. Astfel utilizatorii pot scrie aproape orice atât în câmpurile referitoare la skill-uri, companii la care au lucrat, universități la care au studiat sau chiar specializări. Pentru a rezolva acest tip de problemă este necesară investigarea manuală a datelor împreună cu algoritmi de dezambiguizare a textelor oferite de utilizatori.

Odată extrase informațiile putem extrage șabloane referitoare la modul în care utilizatorii își schimbă joburile, la legăturile între carieră și educație, la legăturile între anumite universități/facultăți și tipul de carieră pe care îl urmează absolvenții acestora.

MODELUL CONCEPTUAL

După cum menționam în introducere acest studiu îl realizăm cu scopul de a realiza o aplicație destinată recrutării în domeniul IT. Modelul utilizatorului pe care îl vom realiza va trebui să ia în considerare acest aspect. De ce afirmăm acest lucru? Pentru că spre deosebire de aplicațiile de gestiune a personalului în care erau importante competențele companiei și modul în care utilizatorul se mulează pe acestea și cum poate fi ajutat utilizatorul să-și dezvolte competențele, aici pe lângă potrivirea între competențele utilizatorului și descrierea jobului trebuie să ținem cont de dorințele și intențiile

utilizatorului. Această idee este nouă, nefiind menționată în aplicații similare [1]. Motivul pentru care dorințele nu au fost luate în considerare au fost că recrutarea a fost privită dintr-un complet alt punct de vedere. Modelul abordat a fost cel în care persoana interesată de un job trimitea un CV și o scrisoare de intenție și rolul software-ului era să determine ce CV-uri se potriveau cel mai mult jobului curent. Recrutarea s-a schimbat în era Web 2.0. Dacă înainte așteptam ca oamenii să aplice la job acum există companii de head-hunting care trebuie să identifice oamenii potriviți pentru un job dat și care în plus și-ar dori să:

- Părăsească firma unde lucrează acum
- S-ar potrivi cu noua firmă
- Și-ar dori să urmeze direcția în carieră dată de noul job
- Este probabil să își schimbe jobul în viitorul apropiat.

Toate aceste informații trebuie să se regăsească în modelul utilizatorului sau să poată fi obținute algoritmic pe baza acestui model.

Modelul utilizatorului așa cum l-am conceput va fi descris prin următoarele propoziții:

- Job=(HSSet,SSSet) (1)
 - Skill=(SkillName,Level) (2)
 - Intentions=(SkillDep, Fidelity, Direction, Slope) (3)
 - PersonProfile=(HSSet,SSSet,Intentions) (4)
- (1) Afirmă că un job este definit printr-un set de skill-uri hard (din domeniul dat) și un set de skill-uri soft (generale). Unim astfel cunoștințele și know-how-ul într-un singur termen – skill – deoarece în sursele de date pe care le deținem acestea nu sunt diferențiate.
 - (2) Un skill este definit printr-un nume și un nivel
 - (3) Intențiile sunt definite prin SkillDep – dependența între o persoană și un skill (este capabil să accepte un job mai puțin legat de cunoștințele sale actuale – de exemplu pentru un programator care a făcut numai C-ar accepta un job pe Java?), Fidelity – fidelitatea față de firmă (cât de des schimbă locul de muncă), Direction – încotro se îndreaptă utilizatorul (spre joburi pur tehnice și în ce zonă a domeniului, spre joburi de management, spre joburi interdisciplinare), Slope – exprimă viteza de avansare a utilizatorului (unii oameni sunt foarte capabili și avansează în poziții superioare mai repede decât ceilalți)
 - (4) Această propoziție afirmă că un profil personal este constituit dintr-un set de skill-uri dependente de domeniu (hard-skills) dintr-un set de skill-uri generale (soft-skills) și din ansamblul de intenții ale persoanei respective

Pe baza propozițiilor (1) și (4) putem determina potrivirea dintre o persoană și un job în felul următor (exemplu simplificat pentru ușurința urmăririi):

- Job={Java,5}, {J2EE,3}

- $Person = \{(C\#,4)\}$
- Presupunem pentru acest exemplu că în ontologia domeniului similaritatea între conceptele Java și C# este 0.8 și între J2EE și C# este 0.3
- $Similaritate(Person, Job) = \{(Java, 0.8*5), (J2EE, 0.3*3)\}$
- $Potrivire(Person, Job) = Job \bullet Similaritate(Person, Job)$

DATE PRELIMINARE

Pentru realizarea acestei aplicații vom utiliza o ontologie a domeniului descrisă extins în [9]. Ontologia este realizată semi-automat pe baza unui corpus de 50000 de descrieri de job-uri din domeniul IT, colectate de pe site-uri de anunțuri de joburi din USA.

Pasul al doilea a fost studierea unui număr suficient de mare de CV-uri extrase de pe LinkedIn. În acest scop am apelat la colegii din grupul de cercetare care au completat un formular⁵ prin care ne ofereau un token care să ne permită să efectuăm interogări pe LinkedIn din partea lor. Pe baza celor 6 tokeni obținuți am extras de pe LinkedIn 2492 de profile complete. Din datele obținute nu putem preciza naționalitatea participanților dar putem vedea în ce țări se află acum.

nr CV	Țară
1864	Romania
105	United States
61	Netherlands
59	United Kingdom
45	France
43	Germany
37	Switzerland
15	Canada
10	Belgium
9	Czech Republic
9	Italy

Tabel 1: Distribuția pe țări a utilizatorilor din baza de date

După cum vedea cei mai mulți participanți se află în România, un număr foarte mare (300-350) aflându-se în Uniunea Europeană.

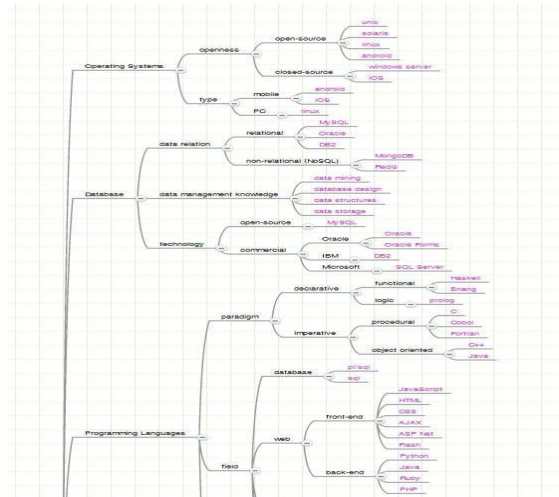


Figura 1 Parte a clasificării skill-urilor tehnice din ontologia domeniului IT

Pentru toate aceste persoane am examinat CV-urile și am analizat skill-urile și am încercat să dezambiguizăm denumirile din datele referitoare la educație.

Analiza manuală a datelor arată că în cele aproape 2500 de profile am descoperit 185 de moduri distincte în care utilizatorii LinkedIn au afirmat că au efectuat un bachelor. Aceste moduri includeau toate variantele de scriere a termenului bachelor, toate prescurtarile posibile precum și includerea în termen/prescurtare a altor calificări obținute.

Un număr ceva mai mic de lexicalizări a fost descoperit pentru studii de master, de PhD sau alte tipuri de studii terminate cu certificări.

Problema s-a dovedit și mai dificilă pentru identificarea instituțiilor de învățământ. Pornind de la 1764 de instituții de învățământ menționate în cele 2500 de CV-uri am reușit să identificăm 89 de instituții de învățământ superior distincte, unele din ele (ca de exemplu Universitatea Politehnica din București) având peste 40 de lexicalizări diferite ale numelui. Tabelul următor prezintă principalele universități identificate împreună cu numărul de menționări în profilele utilizatorilor

Nr CV	Universitate
1513	UPB
261	ASE
163	Univ. Bucuresti
56	Univ. Alexandru Ioan Cuza Iasi
47	Universitatea Craiova
42	Univ. Babes-Bolyai
32	Univ. Tehnica Iași
23	Vrije Amsterdam
20	Univ Spiru Haret
18	SNSPA
12	Caltech
11	Univ. Romano-Americana
11	Univ. Politehnică Timișoara

⁵ <http://itresearch.webfactional.com/website/contribute>

10	Acad. Tehnica Militara Bucuresti
10	Univ. Utrecht
10	EPFL

Tabel 2: Distribuția pe universități a absolvenților din baza de date

Pentru a reprezenta vizual skill-urile utilizatorilor am folosit următoarea abordare. Am identificat skill-urile distincte menționate pe LinkedIn, am verificat de câte ori apare fiecare în setul de CV-uri și am generat un cloud tag folosind valoarea logaritmului numărului de apariții al skill-ului ca pondere în cloud tag. Rezultatul îl puteți urmări în figura 2.



Figura 2 Cloud tag conținând skill-urile celor 2500 de utilizatori de linkedin

CONCLUZII ȘI PAȘI DE URMAT

Lucrarea descrie primii pași spre realizarea unei aplicații semantice de recrutare. Este descris un model al utilizatorului și inițiată realizarea unei ontologii și a unei baze de cunoștințe referitoare la utilizatori. Sunt prezentate date analizate referitoare la profilele acestor utilizatori, date ce vor fi folosite pentru a extrage șabloane despre evoluția în carieră a profesioniștilor din domeniul IT, pentru a le înțelege acestora intențiile.

Cercetarea folosită poate fi folosită atât pentru recrutare cât și pentru realizarea unei comunități de alumni sau pentru un profil tehnologic al companiilor din România. Pentru a ilustra afirmația de mai sus încheiem concluziile cu un grafic realizat pe baza datelor extrase din LinkedIn și care arată repartizarea pe glob a absolvenților Universității Politehnice din București.



Figura 3 Vizualizare a repartizării pe glob a studenților absolvenți ai Politehnicii București

REFERINȚE

1. Mochol et al. - Practical Guidelines for Building Semantic eRecruitment Applications (ICKM06) <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.91.8052&rep=rep1&type=pdf>
2. Radevski, Trichet - Ontology-Based Systems Dedicated to Human Resources Management: An Application in e-Recruitment (OTM06) <https://www-vs.informatik.uni-ulm.de/DE/intra/bib/2006/OTM/papers/4278/42781068.pdf>
3. Mochol et al. - Improving the accuracy of job search with semantic techniques - <http://wissensnetze.ag-nbi.de/publ/BIS2007.pdf>
4. Braun et. al. People Tagging & Ontology Maturing: Towards Collaborative Competence Management http://publications.andreas.schmidt.name/Braun_Kunzmann_Schmidt_PeopleTagging_book_2009.pdf
5. Posea, V. and Harzallah, M., Building a competence ontology. In: Proceedings of the Workshop Enterprise Modelling and Ontology of the International Conference on Practical Aspects of Knowledge Management (PAKM 2004)
6. Paquette G., An Ontology and a Software Framework for Competency Modeling and Management http://www.ifets.info/journals/10_3/ets_10_3.pdf#page=6
7. Schmidt A, Kuntzmann C., Towards a Human Resource Development Ontology for Combining Competence Management and Technology-Enhanced Workplace Learning <http://www.mendeley.com/research/towards-human-resource-development-ontology-combining-competence-management-technologyenhanced-workplace-learning-1/>
8. Berio et al, Knowledge management for competence management - http://i-know.know-center.tugraz.at/content/download/392/1553/file/Berio_paper.pdf
9. Berteanu V., Motor de căutare pentru locuri de muncă, lucrare de diplomă 2011, coordonator Vlad Posea