11-2018

JobSense: A data-driven career knowledge exploration framework and system

Xavier Jayaraj Siddarth ASHOK
Singapore Management University, xaviera@smu.edu.sg

Ee-peng LIM
Singapore Management University, eplim@smu.edu.sg

Philips Kokoh PRASETYO
Singapore Management University, pprasetyo@smu.edu.sg

DOI: https://doi.org/10.1109/ICDMW.2018.00200

Citation

ASHOK, Xavier Jayaraj Siddarth; LIM, Ee-peng; and PRASETYO, Philips Kokoh. JobSense: A data-driven career knowledge exploration framework and system. (2018). *2018 IEEE International Conference on Data Mining Workshops (ICDMW): Singapore, November 17-20: Proceedings*. 1-6. Research Collection School Of Information Systems.
Available at: https://ink.library.smu.edu.sg/sis_research/4261

JobSense: A Data-Driven Career Knowledge Exploration Framework and System

Xavier Jayaraj Siddarth Ashok
Living Analytics Research Centre
Singapore Management University
Singapore
xaviera@smu.edu.sg

Ee-Peng Lim
Living Analytics Research Centre
Singapore Management University
Singapore
eplim@smu.edu.sg

Philips Kokoh Prasetyo
Living Analytics Research Centre
Singapore Management University
Singapore
pprasetyo@smu.edu.sg

***Abstract*—Today's job market sees rapid changes due to technology and business model disruptions. To fully tap on one's potential in career development, one has to acquire job and skill knowledge through working on different jobs. Another approach is to seek consultation with career coaches who are trained to offer career advice in various industry sectors. The above two approaches, nevertheless, suffer from several shortcomings. The on-the-job career development approach is highly inefficient for today's fast changing job market. The latter career coach assisted approach could help to speed up knowledge acquisition but it relies on expertise of career coaches but experienced career coaches are scarce, and they too require update of jobs and skills knowledge.**

Meanwhile, with wide adoption of Online Professional Networks (OPNs) such as LinkedIn, Xing and others, publicly shared user profiles have become a treasure trove of job and skill related data. Job and skill related information is also hidden in the sea of online job posts and ads. Manually exploring and acquiring knowledge from these varieties of information are daunting and time-consuming. On the other hand, one needs substantial effort to personalize the acquired knowledge to his/her career interests. There is a dire need for a self-help tool to ease this knowledge acquisition and exploration problems. Before that, there is also a need to create and maintain a large knowledge base of these jobs, skills and careers. Our data-driven, automated knowledge acquisition and interactive exploration system, JobSense, would help users meet the above challenges. JobSense enables users at several stages of career, to explore this knowledge at ease via interactive search, easy navigation, bookmarking of information entities and personalized suggestions. Also we have introduced a career path generation module, to return relevant career paths to the users.

I. Introduction

Motivation. With rapid changes in the job market and business models, job seekers and career experts are facing challenges staying abreast with the latest job and skill knowledge [1]. Experts from market research firms have mentioned that many jobs are fast becoming obsolete due to this pace of advancements [2]. Traditionally, job seekers can seek knowledge about the jobs, skills and career plans by consulting trained career coaches. Career coaches are however limited by number. At the same time, even career coaches find it hard to keep up with the scale and velocity of the job market changes.

There are efforts to create and update jobs and skills knowledge base for job seekers, career coaches and other stakeholders. For example, European Skills/Competencies qualifications, and Occupations (ESCO) is a knowledge framework in which job titles are associated with skills/competencies and qualifications. Creation of job related knowledge bases such as ESCO is often based purely on manual efforts making them not scalable [3].

There is an increasing adoption of digital platforms for various job related activities such as users on Online Professional Networks (OPNs) sharing their job profiles, and employers using job portals and OPNs to post jobs. These platforms provide much data that enable a scalable and data-driven approach to automate jobs and skills knowledge acquisition. By analyzing these data, one can acquire knowledge and insights directly useful to job search and career guidance.

Nevertheless, this data-driven approach still faces the following challenges:

- As the number of platforms are vast and varied, their data are not standardized by format and by semantics.
- It is difficult to understand the relevance of job or skill entities to one another, as well as other relationships among them. For example, a skill may be similar to another skill, a skill may be relevant to a job, some jobs may be relevant for an industry, etc.
- It is unclear how one can automate job recommendation, skill recommendation, and career planning using the data, leverage on external career knowledge to enhance the career knowledge of job seekers, and address the cold start problems for new users.
- It is not easy to get job and skill-related knowledge to the hand of general public members.

Objectives. Considering the above challenges, we have proposed and built a data-driven mobile web application called **JobSense** to automatically acquire job and skill-related knowledge using information retrieval and deep learning techniques. Our goal is to help job seekers and career coaches explore this knowledge base at ease with advanced user friendly features. JobSense leverages on large amount of data available from OPNs and job portals. The conversion from raw data to graph representation knowledge require a combination of machine learning and information retrieval techniques. Finally, JobSense is designed to have an interactive interface to support simple knowledge navigation.

Contributions. The main contributions of JobSense include:

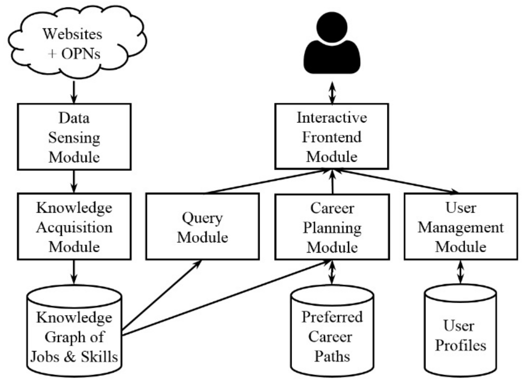

Fig. 1. JobSense's System Framework

- Gathering and integrating jobs, skills and careers data from multiple data sources. This ensures that JobSense can reach out diverse set of job titles and skills;
- User-friendly knowledge exploration via interactive search and navigation on the learned jobs and skills knowledge;
- Supporting popular career path suggestion to progress in a career; and
- Handling of personalized jobs/skills suggestions based on bookmarked skills and jobs.

II. Overview of System Framework

We design JobSense to be an interactive knowledge acquisition and exploration tool to serve the following groups of users:

- *Students and Job Seekers*: They have yet to join the job market, or are looking for jobs. They need to determine jobs that they are interested in, the associated job titles, and skills required to secure the jobs.
- *Working professionals*: They are likely to be interested in future career plans, finding pathways to move up career ladder, or switching career tracks. They also need to identify skills to pick up before applying for their target jobs.
- *Career Coaches*: They are responsible for helping others find jobs and plan career in the fast evolving job market. They need to stay abreast with the wide range of job titles and skills.

With the above target users in mind, JobSense is designed based on the system framework as shown in Figure 1. The framework consists of several components, namely:

- **Data sensing module**: This module ingests data from multiple job related websites and OPNs so as to provide the data required for other data science components, i.e., knowledge acquisition and career planning modules. Different crawlers are developed to support sensing of these different data sources. Table I summarizes the statistics of collected data.
- **Knowledge acquisition module**: We learn a knowledge graph of jobs and skills to distill the different relationships among them. In particular, jobs are related to one another through skill similarity based on Predictive Text Embedding (PTE) [4], as well as transitions made by people holding these jobs. Skills can also be similar to other skills. In Section III, we shall elaborate the approach we have taken to acquire such knowledge. The knowledge graph learnt will be stored for future queries (by the Query module) and career plan generation (by the Career Planning module). The statistics about the knowledge graph entities are summarized in Table II.
- **Query module**: As the jobs and skill knowledge graph is both large and complex, it is not feasible for a user to explore the entire graph at one goal. The Query module is hence responsible for the retrieval of required subsets of knowledge graph as the user explore different parts of knowledge graph. In the JobSense design, we support *neighborhood queries* which are queries to return the neighboring nodes (either jobs or skills) of a given job (or skill) and neighbor type.
- **Career Planning module**: The module specifically performs optimization of career plans using the job transitions previously performed by users in OPNs. Clearly, not all observed job transitions suit a user's career plan. For example, a job transition to a less senior position is usually not preferred by users at the early stage of their career. The Career Planning module therefore has to generate career plans that adapt to the user's preferred criteria. The generated career plans are stored for future user retrieval. We shall elaborate more about this module in Section IV
- **User Management module**: JobSense maintains for each user a profile that records the user's current job title, skill set, and career preferences. The user management module thus covers user registration, user login, user profile & bookmark maintenance, and user profile queries.
- **Interactive Frontend module**: This module provides users an iterative experience of JobSense based on a mobile interface. Users can explore knowledge based on their needs, explore career plans, bookmark interesting jobs and skills, and perform queries. We will describe the frontend interface of JobSense in Section V.

To implement JobSense, we utilize the following software packages.

- The Knowledge Graph of Jobs and Skills is stored as JSON documents using our Elasticsearch [1] cluster. We use Elasticsearch's distributed shard and full-text search capabilities to cope with large volume of mixed structured and unstructured data.
- The Query module supports a combination of native APIs and remotely accessible APIs. These APIs support REST API calls. With the remotely accessible APIs, JobSense knowledge can even be accessible by related applications[2]. To reduce query latency so as to improve

[1]https://www.elastic.co/products/elasticsearch

[2]Interested readers can email the authors for API access.

TABLE I
STATISTICS OF THE DATASET

Statistics	Count
Number of OPN profiles	4,561,881
Number of job posts	863,375

TABLE II
STATISTICS OF JOBSENSE'S KNOWLEDGE GRAPH

Statistics	Count
Number of distinct jobs (min_sup $\geq$ 5)	86,439
Number of distinct jobs with at least 5 relevant skills	65,200
Number of distinct jobs with similar jobs (min_sup $\geq$ 5)	26,499
Number of distinct jobs having a career path	13,483
Number of skills with similar skills (min_sup $\geq$ 5)	21,681
Number of industry labels	17

user experience, we cache frequently requested or computational costly results in a fast-access server side cache implemented using Redis[3].

- The User Management module has its user login implemented using LinkedIn account authentication. As more users maintain their professional profiles on LinkedIn, it is quite reasonable for users to authenticate themselves using their LinkedIn accounts.
- The User Profile data is stored in MySQL, which includes user sessions, biographies and bookmarks.

III. JOBS AND SKILLS KNOWLEDGE CONSTRUCTION

A job in our context represents a job title in some industry, e.g., "Software Developer" in the "Internet" industry. In JobSense, we maintain a master directory of industry labels and every job title is assumed to be assigned one or more industry label. A mapping table is constructed for each data source to store mapping of the industry labels. Other than industry label heterogeneity which can be handled with mapping tables, JobSense addresses job title heterogeneity with job title normalization by parsing each job title into function, domain and position elements as follows:

- **Function element:** This refers to the term in job title which suggests the main job role, e.g., "manager" in the job title "deputy research manager". Every job title must have a function element.
- **Domain element:** This refers to the job domain, e.g., "research" in the job title "deputy research manager". This element is optional. For example, the job title "senior director" does not have any domain element.
- **Position element:** This refers to the seniority level of the job, e.g., "deputy" in the job title "deputy research manager". The position element is also optional, as some job titles do not cover this element, e.g., "business analyst".

At the end of job title normalization, heterogeneous job titles such as "deputy research manager", "deputy manager of research", and "deputy manager, research" will share the identical normalized job title. This is achieved by conducting lexical and syntactic analysis on the job titles using a dictionary of job elements and rules governing the different ways job titles are constructed from their elements. This way, we significantly reduce job title variations so as to learn knowledge about them more efficiently.

Subsequently, JobSense performs the following steps to extract jobs and skills knowledge graph from the job posts and OPN user profiles.

1) **Job nodes:** The set of normalized job titles form the job nodes in the knowledge graph. We further remove job titles which occur in less than five user profiles and less than five job posts. This will reduce rarely used job titles as they are not likely to be associated with sufficient skill description in the data sources. Furthermore, they are also unlikely to be the main target for career exploration.
2) **Skill nodes:** While job titles can be easily found in job posts and OPN user profiles, skills can be mentioned in different ways and more embedded in the description of job posts. In JobSense, we therefore first construct a skill dictionary from extracting skills *explicitly mentioned* by a large set of OPN users. After removing skills that appear in less than five user profiles, we obtain the dictionary of skills, or equivalently the skill nodes of our knowledge graph.
3) **Relevant-Skill-of-Job relationships:** We assume that the same skill dictionary is shared by job posts and OPN user profiles. Given a job title, we determine its top-k relevant skills by representing each job title as a document with skills gathered from all job posts with the job title and all user profiles with the job title included as part of job history. We then rank skills by their TF$\times$IDF score [5] where TF and IDF represent the term frequency of a skill in the job title document and the inverse frequency of the skill among all job title documents. To avoid irrelevant skills being wrongly associated with a job, we set $k = 20$.
4) **Similar-Job relationships:** For a given job, the top k similar job titles are determined by a network embedding technique called PTE [4], applied on job titles with context defined by their associated skills. Once job titles are mapped to the new embedding space, the most similar job titles can be located in the neighborhood of the target job title. For simplicity, we restrict the similar job titles to share the same industry as the target job title. Here, we set $k = 10$.
5) **Similar-Skill relationships:** Like Similar-Job relationships, we apply PTE on skills with their context defined by job titles containing them. Finally, the top k similar skills for a given skill can be found in the embedding space. Here k is set to 10.
6) **Job transitions:** A job transition is extracted from some OPN user profile when the user moves from one job to another job. Note that job posts do not carry job transition information and thus are not used here.

[3]https://redis.io/

IV. CAREER PATH GENERATION

Career path generation is a unique functional feature of JobSense. It is designed to return *relevant* and *useful* career paths to users. A career path here is defined as a sequence of job titles starting from a *source job title* which could be the current job of a user, or any job the user is interested in. A relevant and useful career path thus can be decomposed into one or more relevant and useful job transitions such that each job transition (except the one involving the last job title), say $j_a \rightarrow j_b$ is followed by another job transition $j_b \rightarrow j_c$.

It is non-trivial to determine whether a job transition is relevant and useful to a user. In JobSense, we develop three different approaches:

- **Popular job transitions**: The popularity of a job transition is defined by the number of people adopting the transition in their OPN user profiles. More popular job transitions are thus more relevant and useful. We remove job transitions that involve only one user only as they are rare.
- **Heuristic job transitions**: There are heuristics rules that can generate relevant and useful job transitions based on observed job titles only. For example, even when no job transition is observed between two job titles "data scientist" and "senior data scientist", one can generate the transition, "data scientist" $\rightarrow$ "senior data scientist", as the latter contains a more senior position element in the job title. With the ability to normalize job titles and to extract job title position elements, we incorporate a set of such heuristics rules to generate unobserved job transitions (over observed job titles) that can be used to construct interesting career paths.
- **Uptrend job transitions**: While popular job transitions and heuristic job transitions are likely interesting to users, the remaining job transitions observed from user job histories may not. Some observed job transitions could suggest job demotions or reduction of salaries. To utilize the remaining job transitions meaningfully, we seek to assign *job level* to each job title by averaging the number of years taken to reach the job title since the university graduation year for all users who have ever taken the job title [6]. A job transition $j_a \rightarrow j_b$ can therefore be relevant and useful if j_b has a higher job level than j_a. With more observed job attributes (e.g., salary, company size), we can derived more uptrend job transitions based on criteria defined on these job attributes.

Next, we need to generate career paths by combining the above identified job transitions. For a given source job title, multiple career paths may be constructed. In JobSense, we may rank them based on criteria desired by users. In this demo, we shall just present these career paths altogether.

V. INTERACTIVE USER INTERFACE

In this section, we give a quick walkthrough of JobSense user interface.

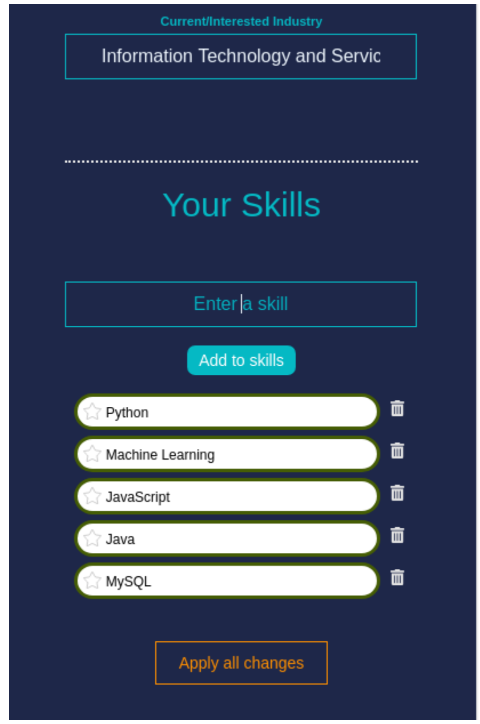

Fig. 2. User Profile Maintenance

1) **User Profile Maintenance:** Figure 2 depicts the mobile user interface for user to update her profile information which include current job title, industry and skills. To minimize input errors and to stick to the dictionaries of industry labels, job titles and skills, the input boxes have populated with values from the corresponding dictionaries and the lists of values will be interactively refined as the user starts entering parts of the input value.

2) **User Dashboard:** As shown in Figure 3, a user is able to view a dashboard of her information. The dashboard shows the user's skills and bookmarks of different interested entities including job titles, skills, and industries.
3) **Search and Exploration Interface:** The search interface of JobSense can be used to query job titles, skills and industries. The search is constrained to the corresponding job title, skill and industry dictionaries. For job title search, the information page of the selected job title will be shown to the user. As shown in Figure 4, the information page of "Software Engineer" job title from the Financial Services industry shows other jobs similar to "Software Engineer", relevant skills (see Figure 5), career plans that originate from the "Software Engineer" job title.
As shown in Figure 6, a career plan starting from the "Software Engineer" job suggests the user to gain more experience to become a "Senior Software Engineer" and subsequently a "Vice President " by switching to management based job roles.

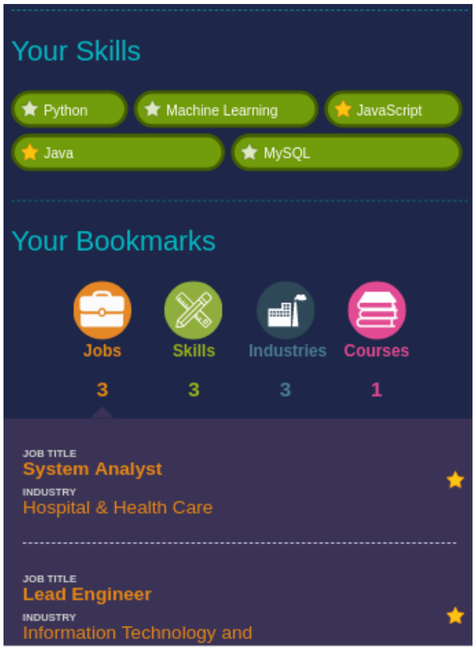

Fig. 3. User Dashboard

Fig. 4. Information Page of similar jobs to "Software Engineer"

4) **Personalized Recommendation:** Based on the user's bookmarked skills and jobs, JobSense recommends other jobs and skills for user consideration. For example, in Figure 7, the user has bookmarked skills such as Statistical Modeling, Data Mining and Statistics, receiving job recommendation aligning to Data Science and R&D such as Data Scientist, Research Scientist etc.

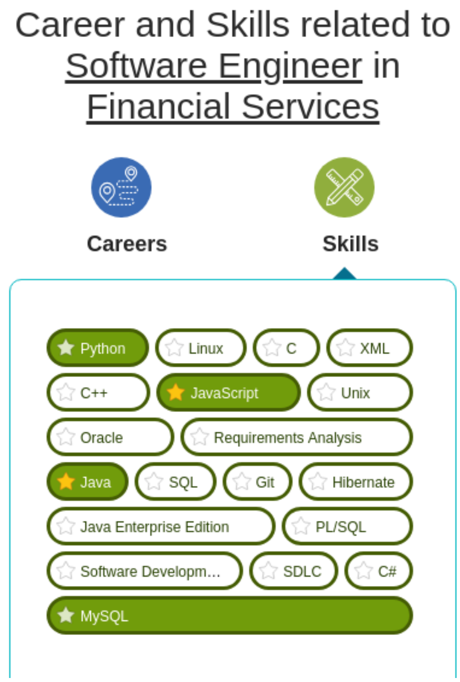

Fig. 5. Information Page of relevant skills to "Software Engineer"

In the case of bookmarked jobs, Figure 3 shows Software Engineering roles such as System Analyst, Lead Engineer etc. for which the user gets recommended skills related to Software Engineering like SQL, Java, SDLC, XML, etc. which is illustrated in Figure 8. These recommendations are derived from the jobs and skills relevant to the bookmarked skills and jobs. To handle the cold start problem for first time users who may not have any bookmarked jobs and skills, JobSense recommends jobs and skills that are popular in their current industry.

VI. Conclusions

Understanding jobs and skills in labor market is not trivial. Jobs and skills information are scattered in job advertisements and online professional profiles on the web and they change rapidly. To automate this knowledge acquisition using a data-driven approach, we introduce JobSense, an unified online platform to learn and explore knowledge about jobs, skills, careers. At present, JobSense is mobile friendly. It automatically generate information pages for jobs, skills and industries for easy browsing. It also caters to user specific interests by maintaining user profiles so as to personalize suggestions of job titles, skills and career plans.

As part of future research, JobSense will be enhanced with user-demand driven acquisition of jobs and skills knowledge so as to offer more relevant and useful information to its users. It will also enrich its current knowledge with new entity attributes so as to support comparisons of knowledge entities. We are also interested to experiment with new deep learning techniques to provide more personalized job/skill/career guidance.

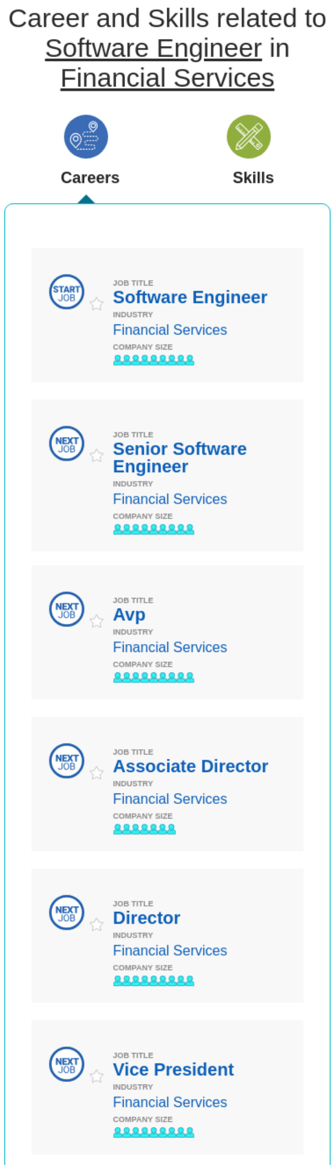

Fig. 6. Information Page of career plan starting with "Software Engineer"

Acknowledgment

This research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its International Research Centres in Singapore Funding Initiative.

References

[1] P. Brown, A. Hesketh, and S. Williams, "Employability in a knowledge-driven economy," *Journal of Education and Work*, vol. 16, no. 2, pp. 107–126, 2003.

[2] C. B. Frey and M. A. Osborne, "The future of employment: How susceptible are jobs to computerisation?" *Technological Forecasting and Social Change*, vol. 114, pp. 254 – 280, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0040162516302244

[3] J. D. Smedt, M. le Vrang, and A. Papantoniou, "ESCO: towards a semantic web for the european labor market," in *Proceedings of the Workshop on Linked Data on the Web, LDOW 2015, co-located with the 24th International World Wide Web Conference (WWW 2015), Florence, Italy, May 19th, 2015.*, 2015.

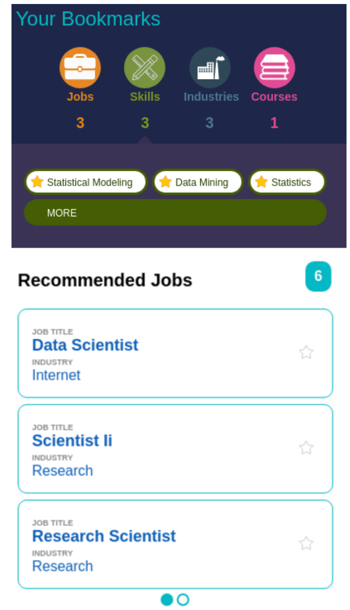

Fig. 7. Job recommendation page for the bookmarked skills

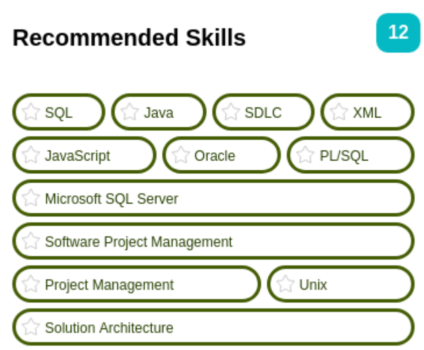

Fig. 8. Skill recommendation page for the bookmarked jobs

[4] J. Tang, M. Qu, and Q. Mei, "Pte: Predictive text embedding through large-scale heterogeneous text networks," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '15. New York, NY, USA: ACM, 2015, pp. 1165–1174. [Online]. Available: http://doi.acm.org/10.1145/2783258.2783307

[5] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. New York, NY, USA: Cambridge University Press, 2008.

[6] R. J. Oentaryo, X. J. S. Ashok, E. Lim, and P. K. Prasetyo, "On analyzing job hop behavior and talent flow networks," in *2017 IEEE International Conference on Data Mining Workshops, ICDM Workshops 2017, New Orleans, LA, USA, November 18-21, 2017*, 2017, pp. 207–214.