

Analyzing User Patterns to Derive Design Guidelines for Job Seeking and Recruiting Website

Yao Lu

École Polytechnique Fédérale de Lausanne (EPFL)
Lausanne, Switzerland
e-mail: yao.lu@epfl.ch

Denis Gillet

École Polytechnique Fédérale de Lausanne (EPFL)
Lausanne, Switzerland
e-mail: denis.gillet@epfl.ch

Sandy El Helou

École Polytechnique Fédérale de Lausanne (EPFL)
Lausanne, Switzerland
e-mail: sandy.elhelou@epfl.ch

Abstract—More and more competitive websites are targeting online job seeking and recruiting. In this paper, we discuss user pattern in an online job seeking website in Switzerland, by analyzing user profiles and actions. Then based on our findings, we derive design guidelines to improve both the interface usability and the efficiency of the embedded jobs/candidates recommender system.

Keywords—patterns; job seeking; recruiting; Web design; recommender system;

I. INTRODUCTION

Nowadays, job seeking and recruiting websites are becoming more and more popular [1]. These websites not only provide a prospective platform for information gathering but also support rich interaction features leveraging user interests [2][3]. From a business point of view, it is crucial to understand user patterns and make use of them, in order to improve the provided services and satisfy the customers [4].

Analyzing user patterns is widely used not only for identifying group related activities but also for user behavior prediction based on previous users [5][6]. Patterns preserve common problem-solving knowledge to give recurring solutions for user-friendly design [7]. Unfortunately, there are only a few studies on online user patterns and design of the job seeking and recruiting websites [8]. In this paper, we analyze user profiles and interaction patterns on a job seeking and recruiting website.

Usually, users of job seeking and recruiting websites split into two main categories: employer and candidates. They participate in the job seeking and recruiting process by completing their profiles and performing different actions. By relying on usage statistics, graphical visualization and analysis, we derive typical user interaction and profile patterns for both user categories. Based on the discovered patterns, we generate design guidelines for jobs/candidates recommender system as well as for the website's interface. The guidelines proposed in this paper have contributed to increasing the number of overall visits of company,

and job profiles by 300% within the period of one month.

The rest of the paper is organized as follows. Section 2 gives a general description of the website's dataset. Section 3 discusses user profile and interaction patterns. Section 4 provides guidelines for designing the interface and the embedded recommender system, based on the analyzed user patterns and the experience with our case study.

II. DATASET INTRODUCTION

The dataset used in our case study includes more than 7,000 candidate users (40% of which have provided CVs), 320 companies, and 7,000 jobs. The time span is from May 2008 to the beginning of February 2012.

Candidates, company, and job-related statistics are given below.

A. Candidate Statistics

- 80% of the candidates live in Switzerland.
- There are twice more male than female candidates.
- The average age is 27, and 88% of the candidates are between 22 and 35.
- 98.5% of the candidates are university students or graduates.

B. Company Statistics

- One in six companies is a startup.
- Most of the companies are located in Switzerland and Germany takes up another big part.
- 320 companies cover 50 industry branches. The biggest industry group is IT, Internet services, and computer hardware/software. The next biggest group is management consulting.

C. Job Statistics

- 89% of the posted jobs are hosted externally and the rest is directly posted on the website (using predefined forms).
- 61.7% of the jobs are located in Switzerland and 37.3% in Germany.

III. USER PROFILE AND INTERACTION PATTERNS

In this section we discuss user profile and interaction patterns in our case study.

A. Input Patterns in User Profiles

Of most job seeking and recruiting websites, user profiles are important for both employers and candidates. Taking candidates for example, they can either input their information online or upload CV files. Meanwhile, many websites support users to import their profile from other websites like LinkedIn [9].

Online profile forms contain specified fields. For candidates, the main profile fields typically consist of personal information (name, gender, birthday, etc.), educational background, and work experience. For companies and jobs, the main fields are location, industry, description, and requirements. Data processing is usually easier when dealing with structured and standardized profile information, as in our case study. Profile input patterns can be summarized as follows:

- Almost all companies and jobs have their basic information filled, and 70% provide more details regarding offers and requirements.
- Considering the candidates who filled their online forms, the profile completion rate is rather satisfying. 86% of these candidates input their personal information (name, gender, age). Moreover, 99.7% of them input their education while only 51.4% input their work experience (probably because most of them are graduating students). The completion rate for the different profile fields is illustrated in Figure 1.
- There is a preference for uploading CVs rather than filling online forms. In our case study, the website was supporting CV upload from November 2009 to October 2011. During this period, 57% of the candidate profiles were uploaded PDF files.

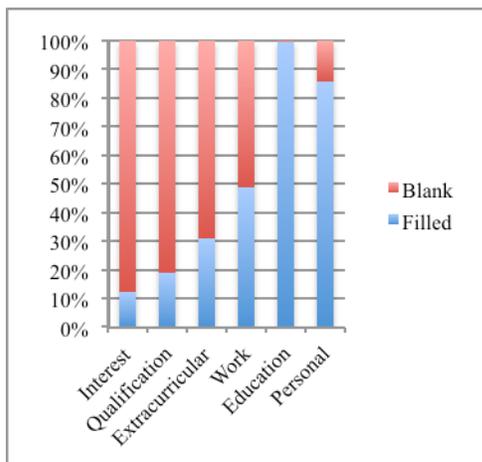


Figure 1. Profile completion rate per field

B. Similarity Patterns in User Profiles

On many websites, there is a recommendation list of items similar to the one a user is visiting [10]. For example, when a candidate is visiting a job page, similar jobs (in terms of location, industry, description) are recommended to that candidate. Similarly, when an employer is on a candidate's page, other candidates with similar profiles (in terms of educational background, skills, work experience) are also recommended to that employer. Figure 2 shows an example of similar job recommendation on LinkedIn [9]. A prerequisite to provide such features is to compute similarity between entities of the same type.

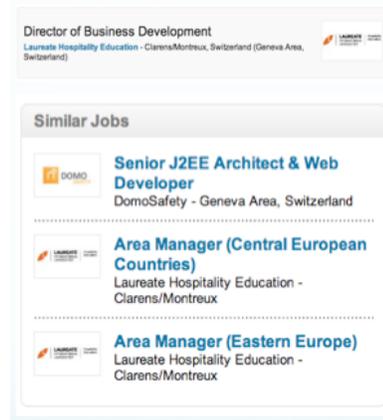


Figure 2. Similar job recommendation in LinkedIn^[9]

In this section, we discuss profile similarity patterns illustrated using similarity graphs. Profile similarity consists of measuring the extent to which two user profiles are similar in terms of content. We construct similarity graphs for all types of users. In these graphs, an edge connects two user profiles if they have a similarity higher than a minimum threshold. We discovered that the graph of candidates and employers follow different patterns. We use an information visualization tool named aiSee [11] to build and visualize graphs of job, candidate, and company similarity. The profile similarity measurement is shown in Figure 3.

As it is illustrated in Figure 3, there are two types of profiles: structured and unstructured. Structured profiles consist of several predefined fields that can be filled online like name, age, and education. Unstructured profiles, on the other hand, are uploaded with no standardized format (e.g. uploaded CV files). For structured profiles, the overall similarity consists of a linear combination of weighted field similarity in the case of fields having predefined values and normalized content-based similarity in the case of free text fields. Table 1 shows the fields in candidate's profile we select and the weight assigned to each field. The choice of the fields and their corresponding weights are set after discussion with company representatives. For them, the educational background, the university, and the degrees are the most important fields since their target customers are graduating students. When it comes to companies and jobs, the industry field, geographic location, job title/position and its requirements are important fields to consider for

similarity measurement [12]. When it comes to unstructured profiles, it is difficult to extract corresponding fields. So, when comparing two PDF files or one PDF file and one structured profile, both are parsed into unstructured data and LSA (Latent Semantic Analysis) is used to compute their similarity. Both structured and unstructured profiles have textual content (i.e. description fields in online filled forms). In this case, we also use LSA to compute similarity. More specifically, we call APIs provided by Salsadev [13]. Through the process described in Figure 3, we obtain similarity of structured online profiles and unstructured uploaded profiles.

TABLE I. SELECTED FIELDS AND WEIGHTS OF CANDIDATES' ONLINE PROFILES

Field	Weight (%)	Field	Weight (%)
Gender	5	Age	10
University	10	Study Course	20
Diploma	10	Language	5
Work Experience	20	Qualification	10
Extracurricular	10	Total	100

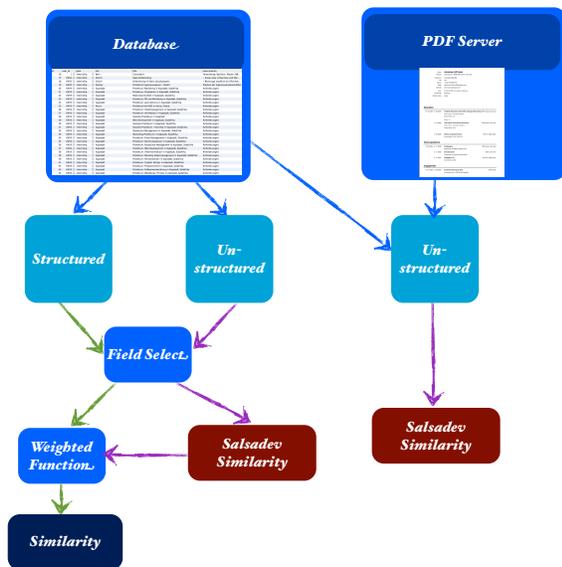
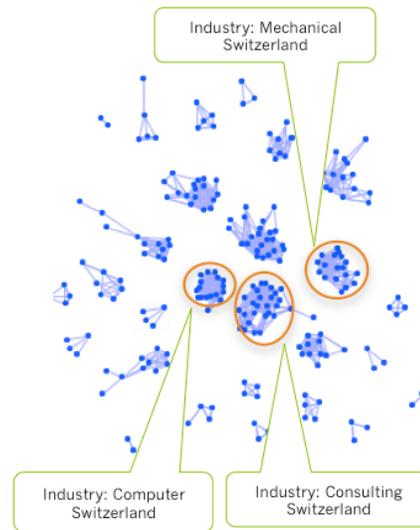
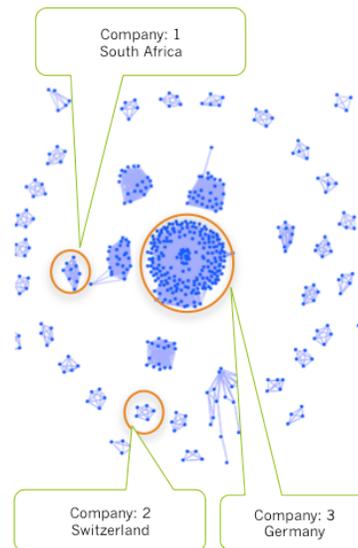


Figure 3. Similarity computation process for candidate profiles

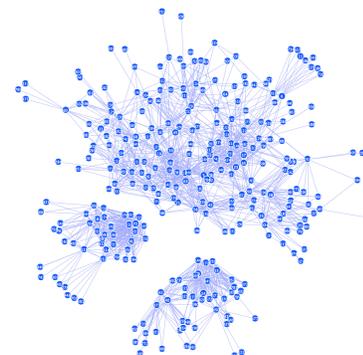
Figure 4 displays the similarity graphs for all user types. In these graphs, nodes represent users, and edges connect similar users. Radial layout is used and users with higher number of incoming edges (i.e. those having more similar users) are located in the center. Ignoring users that have no edges (not similar to any other user), we observe that the graph of candidates follow patterns that are different from other user types (jobs or companies). That is, company graph and job graph consist of clusters of small dense graphs, while the candidate graph consists of big radial clusters. However, according to a preliminary evaluation, this difference in patterns does not affect accuracy in any significant way. The use of structured candidates' CVs yields more accurate results than unstructured CVs.



(a) Company similarity graph



(b) Job similarity graph



(c) Candidate similarity graph

Figure 4. User similarity graphs

Reasons for the difference in similarity patterns are discussed hereafter. In the graphs of jobs (Fig. 4a) and companies (Fig. 4b), there are many small, dense graphs with a small number of nodes but most of them are connected to each other. But in candidates' graph, there are big, radial clusters. Splitting the candidates similarity graph into a graph for structured candidate CVs and another graph for unstructured CVs, still displays the same pattern. This indicates that the patterns observed in candidate graphs are not due to dealing with unstructured CVs. On the other hand, a closer look at company and job graphs, shows that company and job nodes tend to cluster by their dominant factor. In our case, the dominant factor of company is industry. That is because we use location, industry, and scale to compute the similarity and, as it is described in section 2, most of the companies are located in Switzerland. As a result, it is the industry field that makes the difference, and contributes the most in forming clusters of similar companies. When it comes to jobs, the dominant factor is the company that posts them. It is not only because the same company usually offers jobs with similar requirements related to the company's industry but also because job descriptions follow the same style and use the same set of terms. As a result, most jobs posted by the same company form one cluster. But when it comes to candidates, it is difficult to label clusters in the profile similarity graph. The main reason is that there is no dominant feature for candidates. There are much more fields considered for candidates than for companies (country, industry, scale) and jobs (company, location, title, description). In addition to the multiple fields contributing to the similarity pattern, many candidates have more than one input for each field. For example, they may have multiple interests, and may have studied in more than one university. Taking the field as an example, in our measurement, if two students have at least one field of study in common, they are considered as similar along this attribute. For example, if student A studies finance and business, student B studies finance and computer science, they are considered as similar in terms of their field of study. So a group of nodes with multiple, overlapping values are clustered. In contrast, companies and jobs usually have one value in most fields.

In addition, candidate similarity is not necessarily transitive. Let us consider a student C that has studied computer science and graphic design. Ignoring all other fields, C is similar to B because they both studied computer science but C is not similar to A. It is also observed that the nodes in the center of a radial graph tend to have multiple backgrounds. This is due to the fact that there are more nodes similar to them along different directions.

Based on the similarity patterns observed for profiles, potential discussions about the user classification, pattern improvement, and contribution to the recommendation will produce more interesting discoveries. We summarize some limitations exclusively relying on content-based analysis to produce recommendations:

- Candidates tend to mention multiple interests without the possibility to specify their order of preference.

- Two candidates with a high similarity in their profile may not attract employers in the same way.
- While computing profile similarity, it is impossible to set the field weights in such a way to satisfy every user, as all users have different priorities. For example, some candidates focus on employers in the same industry with their majors while other candidates are more concerned about the location of their jobs.

To better understand the users, and deliver personalized jobs recommendation, it is deemed important to leverage users interest by examining how they interact with the website. Actions such as rating a job, liking/disliking a company, adding a company event or a candidate profile to one's favorite list are useful indications of interest. In the next section, we discuss the user interaction patterns derived from logged user actions.

C. User Interaction Patterns

Job seeking and recruiting usually provide some social media features like connect, like, share, and recommend to friends. These features do not only help user discover interest and opportunities, but can also be exploited in recommender systems [14]. Figure 5 shows two screenshots from LinkedIn and Xing [15]. When it comes to our case study, the possible actions are summarized below:

- Visit
- Share
- Like/Dislike
- Rating
- Recommend to friends
- Add to favorite list (or bookmark)
- Apply (for a job)

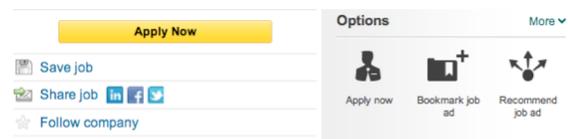


Figure 5. Social media features in LinkedIn and Xing^{[9][15]}

On the interface, most of the buttons to perform interactions and express interest are located in easy-to-use places. According to user action logs, the interactions are sparse and the majority of the users are anonymous. Interactions follow a long tail distribution [16]. In our case, taking visits as an example, 20% of the users contribute to more than 70% of performed visits. Figure 6 shows the long tail by ranking users by the number of their page visits. Meanwhile, majority of users are anonymous. According to data logged between May 2011 and February 2012, on a total of 1.4 million visits, only 2278 were contributed by logged in users, i.e. 98.3% of the visits are anonymous. Interestingly, from April 2011 to September 2011, the website allowed anonymous users to bookmark candidate, job, or company profile pages. The result was that anonymous users were responsible for 99.7% of the actions of bookmarking.

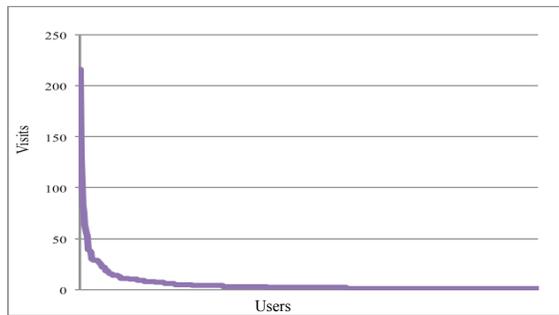


Figure 6. Long tail of performed visits

IV. WEB DESIGN GUIDELINES

In this section, we provide a list of design guidelines based on the case study, and more specifically: the analysis of user profile, interaction patterns and the interface usability analysis. These guidelines can help in designing usable interfaces and job recommendation services for online job seeking and recruiting websites. In our case for example, applying the guidelines provided below, contributed to an increase of 300% in the number of visits within the period of a month:

- Users tend to be anonymous which renders pattern analysis less personalized. So website registration should not be difficult or time consuming. Instead, low entry barriers should be adopted, and profiles could be filled progressively. User interest should be leveraged in a semi-automatic way by combining different techniques such as content-based and interaction-based analysis.
- Mining user interactions on the website complements content-based profile analysis in deriving user patterns and delivering personalized recommendations. Social media features such as bookmarking, liking/disliking and rating should be used to encourage interactions and infer user interest. Nevertheless, when it comes to job recruiting websites, these features should be “built-in” and should not be imported from other sites such as the Facebook “like” button [17] or the Twitter “share” button [18]. The main reason for that is that candidates do not necessarily want that everyone in their social network know that they “liked” a job. Thus, using a Facebook button would discourage them from expressing interest. Last but not least, using external features prevents the website from being able to log the corresponding user actions and exploit them for data mining and recommendation purposes.
- Obviously, the location of the interaction features is essential. Bookmarking or adding to favorites is deemed useful and widely used in many online sites (e.g. social sites, e-commerce, job seeking sites). Surprisingly, we had only very few records of such features in our study. Usability analysis showed that it was the feature’s inadequate position that

dramatically affected its usage. Instead of being placed on the homepage, the user had to go to settings to find his/her list of favorites. So it was time consuming for candidates to perform extra actions to access it in order to bookmark a company or job page.

- Based on observed patterns, users tend to upload CV files rather than fill online forms. Therefore, websites should provide various ways for users to provide their information, like online form, uploading, and file import from other existing sites. On the other hand, since it is easier to process structured data, incentives for filling online forms should be created for users.
- Content-based profile analysis is not by itself sufficient to leverage user interests, while interaction-based analysis has the problem of sparse data, as reported earlier. To overcome each method’s limitations, a hybrid job recommender system, which combines the two techniques, should be adopted. All data types, including structured content, unstructured content, and recorded interactions should be explored. Using hybrid recommendation techniques help cope with the data sparseness and cold start problems.

V. CONCLUSION AND FUTURE WORK

In this paper, we analyze user profile and interaction patterns for an online job seeking and recruiting website. Based on our user pattern analysis, we provided guidelines for the design of the website’s interface and the embedded recommender system. In the future, we will continue to analyze the data from the website for further user pattern analysis. A user study will also be conducted to assess the impact of our design guidelines on the website’s usability and evaluate the accuracy of the embedded job/candidate recommender system.

References

- [1] Jobvite 2011 Social Recruiting Survey. Dec. 2011. <http://recruiting.jobvite.com/>
- [2] P. Lops, M. D. Gemmis, G. Semeraro, F. Narducci and C. Musto, “Leveraging the linkedin social network data for extracting content-based users profiles,” Proceeding of the fifth ACM conference on Recommender Systems (RecSys ‘11), ACM, Oct. 2011, pp. 293-296, doi:10.1145/2043932.2043986
- [3] N. B. Ellison, C. Steinfield, and C. Lampe, “The benefits of Facebook “Friends”: Social Capital and College Students Use of Online Social Network Sites,” *Computer-Mediated Communication*, vol. 12, iss. 4, July. 2007, pp. 1143-1168, doi: 10.1111/j.1083-6101.2007.00367.x.
- [4] Edmond H. Wu, Michael K. Ng, Andy M. Yip, and Tony F. Chan, “A Clustering Model for Mining Evolving Web User Patterns in Data Stream Environment,” Proceeding of Intelligent Data Engineering and Automated Learning (IEDAL 2004), pp. 565-571, doi: 10.1.1.4.8132.
- [5] A. Ross, C. B. Owen, and A. Vailaya, “Models for User Access Patterns on the Web: Semantic Content versus Access History,” Proceedings of 5th Annual World Conference on the WWW and Internet (Webnet 2000), pp. 464-469, doi: 10.1.1.117.7275.

- [6] S. Han, A. Goker, and D. He, "Web User Search Pattern Analysis for Modeling Query Topic Changes," *Information Processing & Management*, vol. 38, iss. 5, 2002, pp. 727-742, doi: 10.1016/S0306-4573(01)00060-7.
- [7] I. Wentzlaff, and M. Specker, "Pattern-based development of user-friendly web applications," *Proceedings of ICWE '06 Workshop proceedings of the sixth international conference on Web engineering*, ACM, July. 2006, doi: 10.1145/1149993.1149996.
- [8] B. J. Jansen, K. J. Jansen, and A. Spink, "Using the web to look for work Implications for online job seeking and recruiting", *Internet Research*, vol. 15, iss.1, 2005, pp. 49-66, doi: 10.1108/10662240510577068.
- [9] <http://www.linkedin.com/>
- [10] B. Sarwar, G. Karypis, J. Konstan and J. Riedl, "Item CF Item-based Collaborative Filtering Recommendation", *Proceedings of the 10th international conference on World Wide Web (WWW '01)*, ACM, May 2001, pp. 285-295, doi: 10.1145/371920.372071.
- [11] <http://www.absint.com/aisee/>
- [12] S. L. Rynes, R. D. Jr. Bretz, and B. A. Gerhart, "The Importance of Recruitment in Job Choice: A Different Way of Looking," *Personnel Psychology*, vo. 44, iss. 3, Sept. 1991, pp. 487-521, doi: 10.1111/j.1744-6570.1991.tb02402.x.
- [13] <http://www.salsadev.com/>
- [14] R. Rafter, Keith. Bradley, and B. Smyth, "Automated collaborative filtering applications for online recruitment services," *Proceedings of the International Conference on Adaptive Hypermedia and Adaptive Web-based Systems*, Springer-Verlag Press, Aug. 2000, pp. 363-368, doi: 10.1007/3-540-44595-1_48.
- [15] <https://www.xing.com/>
- [16] A. Enders, H. Hungenberg, H Denker and S. Mauch "The long tail of social networking. Revenue models of social networking sites", *European Management Journal*, vol. 26, iss. 3, Feb. 2008, pp. 199-211, doi: 10.1016/j.emj.2008.02.002.
- [17] <http://www.facebook.com/>
- [18] <https://twitter.com/>