Assessing Technical Candidates on the Social Web

Andrea Capiluppi, Brunel University

Alexander Serebrenik, Eindhoven University of Technology

Leif Singer, Leibniz Universität Hannover

// Through the Social Web, potential employers can qualitatively evaluate job seekers, even those lacking a formal degree or changing their career path, by assessing candidates' online contributions. //



IN 2009, 48 percent of the Inc. 500 companies had used social media sites for recruitment and candidate evaluation. In a separate survey of 115 smalland medium-sized businesses, Jobvite found that 78 percent of them used social media in their recruiting efforts. The top three sites used were LinkedIn, Facebook, and Jobster.

The traditional recruitment process

is commonly based on the following steps:

1. Summary of qualifications. The first stage of the recruitment process is gathering qualifications from candidates. The aim is to assess a candidate's qualifications against the benchmarks in the job description. Potential employers do this by

- evaluating a candidate's previous experiences (summarized in his or her documentation). Such experiences might include degrees, certificates, courses taken and taught, and previous positions.
- 2. In situ evaluation. A second, optional stage is asking the candidate to sit in on a workbench exercise. This helps potential employers to assess whether the candidate has actually mastered the qualifications listed in his or her curriculum vitae (CV).
- 3. Reference. In the next stage, employers request references from third parties to qualitatively assess the impressions that others received from the candidates.
- 4. *Interview*. In the final stage, a panel of people working for the hiring organization evaluates the candidate in a formal interview.

This procedure has several shortcomings when assessing candidates discovered through social media, especially for those who lack formal degrees, relevant job experience, or IT backgrounds (that is, "nontraditional" candidates).3 In this article, we discuss how the Social Web enables recruiters and developers alike to assess software developers and identify issues that could arise when doing so. In particular, we define a set of metrics and signals that let recruiters and job seekers evaluate both traditional and nontraditional candidates. Employers can see a summary of qualifications and work experience by observing a candidate's online activities in open, repeatable (that is, repeating the same analysis should obtain the same results), explicit, unambiguous terms. Similarly, reference gathering could be supported by monitoring an individual's status or

ABLE 1

Examples and classifications of social network sites.

		Audience					
Purpose		Software developers	Professionals	General audience			
Content sharing		GitHub (http://github.com) bitbucket (http://bitbucket.org) SourceForge (www.sourceforge.net) Google Code (http://code.google.com) Snipplr (www.snipplr.com)		Pinterest (http://pinterest.com) Foursquare (http://foursquare.com) Flickr (www.flickr.com)			
Knowledge sharing	General	Mailing lists Bug trackers	Cross Validated (http:// stats.stackexchange.	Blogs Wikipedia (www.wikipedia.org)			
	Q&A	Stack Overflow (http://stackoverflow.com)	com)	Yahoo Answers (http://answers.yahoo. com)			
	News	Slashdot (http://slashdot.org) Hacker News (http://news.ycombinator. com)		reddit (www.reddit.com) Digg (http://digg.com)			
Professional networking		Careers by Stack Overflow (http://careers. stackoverflow.com)	LinkedIn (www.linkedin. com) Xing (www.xing.com)	Twitter (http://twitter.com) Facebook (www.facebook.com) Google Plus (http://plus.google.com)			
Profile aggregating		Masterbranch (http://masterbranch.com) Coderwall (http://coderwall.com) Geekli.st (http://geekli.st)	Zerply (http://zerply.com)	Yatedo (www.yatedo.com)			
Blogging		Numerous individual blogs spread across multiple sites		WordPress (http://wordpress.org) LiveJournal (www.livejournal.com)			

reputation among his or her peers in the Social Web.

Social Media Sites

Social media have changed how developers collaborate, communicate, learn, and become informed about new technologies.⁴ A vast array of social media sites target software development professionals and amateurs. A superset of social network sites (as defined by Danah Boyd and Nicole Ellison⁵), these social media sites support identity, content, and interaction transparency (as defined by H. Colleen Stuart and colleagues⁶). Table 1 classifies social media sites on the basis of their audiences and purposes.

Social media sites target a wide spectrum of users, ranging from a general audience (for example, Facebook; www.facebook.com) to general professionals (for example, LinkedIn; www.linkedin.com) to IT experts (for example, Stack Overflow; http://stackoverflow.com). Some sites provide the ability to share content such as source code fragments (for example, Snipplr; http://snipplr.com), entire projects (for example, GitHub; http://github.com), or images (for example, Flickr; www.flickr.com). Others support knowledge sharing through questions and answers (for example, Stack Overflow) or news postings (for example, reddit; www.reddit.com).

Some support multiple activities; for example, LaunchPad (http://launchpad.net) supports code sharing, question answering, and communication.

Profile aggregators create a single public user profile from an individual's profiles across various social media sites. We classify these aggregators as either user driven or autonomous. User-driven profile aggregators require a user to provide information, such as the profiles to aggregate. Autonomous profile aggregators actively crawl the Web for any information about an individual without his or her participation or consent.

Masterbranch (http://masterbranch. com) is an example of a user-driven profile aggregator targeted at software developers. Members can import their LinkedIn profiles or Stack Overflow reputations, but Masterbranch won't autonomously gather information about nonmembers. In contrast, Yatedo (www.yatedo.com) crawls the Web to collect all publicly available information on any individuals it can find. Autonomous profile aggregators will likely contain information about more individuals than the user-driven ones. However, user-driven profile aggregators aren't limited to publicly available information as autonomous

ones are; users can provide them with access to private software repositories to crawl. Therefore, these aggregators can provide recruiters with all the information that the developers explicitly choose to share.

Social Media Activity and Content as Signals for Qualifications

The open source development model, although especially technical and requiring specific expertise, has raised the general profile of online user-generated content. At first, it supported a massive collaboration environment; lately, it's also been used to showcase one's skills to peers and recruiters.

Laura Dabbish and her colleagues examined software developers' behavior in public projects on GitHub.⁷ They found that developers consciously manage their online reputations and are aware that other developers judge them by their publicly visible behavior. Developers use those signals to assess both personal characteristics, such as commitment, and work quality. Signals include the type and quality of commitments, the number of forks in a project (the number of project copies that can evolve independently), and the project owners' contribution policies.

One study investigated how software developers use developer aggregators. Developers used simplified signals, such as badges and achievements, to get an overview of another developer's area of expertise. Next, they used more complicated signals, such as posts on Twitter or behavior in online discussions, to derive hypotheses about technical qualifications and soft skills.

Less motivated and self-driven developers—those that one study participant labeled "9-to-5 developers"—were assumed to be less enthusiastic about the work: "A 9-to-5 developer is not tweeting about the latest stuff that's coming out of the W3C mailing

list. A 9-to-5 developer is tweeting a picture of the hamburgers he's frying at 4:30." In other words, this study participant used the subjects of a developer's tweets—in this case, technologies and standards versus leisure activities—as a signal for passion about software development.

Social Media Reputation as a Signal for References

Many recruiters use job references because they can provide insights into candidates' personality traits. Moreover, references originate from colleagues and coworkers who had the opportunity to observe candidates' typical performance, which is a good predictor for how well individuals will perform on average. 10

Social media sites provide various means for members to express appreciation of their peers' achievements. Some sites support highlighting activities within the site itself, such as Stack Overflow's badges and reputation metric. Others allow for the appreciation of activities external to the site-for example, recommendations on LinkedIn are meant to rate past professional collaborations. In both cases, highlighting an achievement provides information about an individual. However, the credibility of such a highlight depends on its meaning among peers—for example, whether the highlight is seen as an actual status symbol. (For further discussion, compare this with "Badges in Social Media: A Social Psychological Perspective."11)

Our discussion is based on several studies that have examined the use of reputation signals on the Social Web. One study found that reputation signals on Wikipedia influence how favorably the community will rate an editor's articles. Another study found that developers often use diverse signals—such as a developer's activity, Twitter and GitHub followers, and the achievements he or she displayed by profile

aggregators—to assess the reputation of other developers. They also actively endorse others in public to appreciate good work, for example, by giving "free beer" on Masterbranch or by clicking the endorse button on Coderwall (http://coderwall.com).

Social Media and Their Signals

Social media sites make various mechanisms available to site members, and there are certain implications of being a contributor to such sites. We can classify the sites according to how distinct and reliable their signals can be. Each site provides a distinctive dataset that could be used to assess a candidate. There are also a number of drawbacks that recruiters and candidates should take into consideration for each site. Table 2 provides an overview of sites and a guide to interpreting them as sources of qualifications and references.

Generic Social Network Sites

Although generic social network sites are mostly used for personal networking, recruiters can use the available signals to infer how an individual might integrate and work in a professional setting.

Qualification signals. Generic social network sites such as Facebook target large, nonspecific audiences. Thus, it's more difficult to crawl them for relevant information about an individual's expertise. However, generic social network sites do help users summarize their past and present personal (although not always professional) affiliations.

References signals. Sites such as Facebook can provide a visible indicator of the type of social network that individuals are able to maintain. They also show how peers acknowledge



Social networks as sources of qualifications and references.

Type of	As a source of qualifications			As a source of references		
social network	Advantages	Disadvantages	Metrics	Advantages	Disadvantages	Metrics
Generic social networking sites	Past affiliations and experience summarized	Blended social and professional profiles	List of past experiences, interests, and passions in posted content	General psychological well-being; reciprocity and volunteerism; peer recognition	Nonreproducibility; partial relation to work performance	Network size; number of followers
Social code- sharing sites	Open, verifiable metrics and signals	Risks of free riding, incorrect attribution, niche marketing, unmet technical expertise requirements, and open source bias	Number of popular repositories authored or contributed*	Indirect measurements by users downloading and using projects	No clear definition of "peers"; technical proficiency required for interpretation	Number of downloads for code; size of developer's community
Q&A sites	Easy to distinguish real experts	Highly skewed between experts and novices; underrepresented minorities	Number of high-quality answers in specific topics; list of topics contributed to	Objectivity; badges and reputation are automatically awarded	Leniency; long- tailed distributions; nonrepresentative- ness	Number of badges associated with self- interest; number of badges associated with prosociality
Professional social networking sites	Could provide dates, duration, position, and team sizes	Duplicated information; nonverifiability of mentioned skills; underrepresentation of nontraditional candidates	List of past jobs, skills, and experience in an area	Closest to traditional references	Origins of recommendations; leniency; visibility of recommendations	Number and nature of recommendations; size of network
Profile aggregators	Signals based on actual code written	Possible oversimplification; technical expertise might be required for interpretation	A combination of all other metrics mentioned	Professional peer recognition	Open source bias	Endorsements received and given; diversity in technologies; activity level; achievements earned through actual code

^{*}A. Capiluppi, A. Serebrenik, and A. Youssef, "Developing an H-Index for OSS Developers," Proc. 9th Working Conf. Mining Software Repositories (MSR 12), IEEE CS, 2012, pp. 251–254.

their activities. In general, participation in such sites has been shown to be beneficial because it provides access to nonredundant information (such as employment connections). From the candidates' perspective, participation in these sites can significantly increase their social skills.¹³ This especially proves beneficial for low-esteem candidates.

Recommendations. Using Facebook (or similar sites) to seek information about potential candidates can be problematic for both candidates and employers. Candidates might have published professionally unattractive, private content, making them less appealing to employers. Companies taking this seriously might judge individuals on materials irrelevant to their profession,

rejecting otherwise perfectly capable candidates. So, even though using generic social network sites for assessing developers has some value, we don't recommend using them for recruiting.

Social Code Sharing

Social code-sharing sites combine software repositories with developer networks. The available signals relate to both the software projects and the interactions among developers.

Qualification signals. Open source projects' contributors and core developers tend to archive and log all development processes in publicly accessible repositories. Social code-sharing sites, such as GitHub, and generic code repositories, such as Google Code (http://code. google.com), present opportunities for communicating candidates' skills and experiences in distributed development teams. Operationalizing an interpretation of these experiences is a matter of summarizing them in a concise, comprehensible, correct way. Metrics and signals about online activity should therefore be open, verifiable, and reproducible.

References signals. Sites that let members host source code typically limit the ways in which reputation can be expressed. They mostly focus on the relationships between individuals and their code rather than between members. SourceForge (www.sourceforge.net) and Google Code are good examples of this model.

On the other hand, GitHub lets users follow one another. To a degree, this reflects professional reputation—users who are doing work that's interesting to their peers will tend to have more followers.

Recommendations. Social code-sharing sites can only indirectly provide information on how an individual achieves his or her reputation—to assess actual reputation, recruiters would need to consider data from all of the individual's peers. We feel that recruiters who aren't familiar with the software development culture should restrict their use of social code-sharing sites mainly to assessing an individual's qualifications. Because the content on those sites (source code, tickets, commits, and so on) is fairly

technical, we suggest that recruiters should seek the assistance of software developers to interpret these signals.

Question and Answer Sites

Q&A sites such as Stack Overflow and Yahoo Answers let users earn badges and rewards, which are public metrics for what members have earned through their activities. These metrics reflect the quality and the types of interactions among individuals. For example, contributors on Stack Overflow can earn the Great Answer badge for high-quality answers, whereas Yahoo Answers users can earn the Best Answer indication for good answers.

Qualification signals. The sites clearly define the areas to which a member has contributed. Metrics such as the number of answers given for a specific topic, the frequency of participation, and the quality of answers as rated by peers make it easy to distinguish a real expert from an occasional contributor.

References signals. Awarding badges and rewards is a very visible approach to assigning reputations to individuals. This practice categorizes people, their skills, and their expertise levels, as well as their peers' perceptions of these aspects. Most statuses and badges are awarded automatically according to a transparent set of rules and can be easily scrutinized.

Recommendations. Recruiters should be careful when evaluating a candidate based on badges and reputations because users tend to up-vote one another's questions, answers, and comments (called *leniency bias*). On the other hand, the distribution of experts and novices is highly skewed: a few quick, knowledgeable individuals earn most of the good ratings, whereas the majority of users earn far fewer.¹⁴

Q&A sites present a real advantage when monitoring contributors' status

and expertise. Scattered evidence suggests that high reputation is immediately visible to recruiters. Whether or not reputation affects job recruiters is subject to debate. One participant said, "Look at three or four of the highly voted answers [the candidate] wrote. If you've ever hired a programmer in your life, it's obvious those people are all some of the best programmers you could ever hire." On the other hand, another participant said, "It's not a case of direct causation—that is, having a high reputation does not lead to a highpaying job. Being intelligent, articulate, passionate and knowledgeable leads to a high-paying job" (http://programmers. stackexchange.com/questions/ 20407/will-high-reputation-in-stack -overflow-help-to-get-a-good-job).

Professional Social Networking Sites

Sites aimed at professionals, such as LinkedIn, let individuals track and publish their career paths, skills and past experiences, the size and tenure of the teams with whom they've worked, and the roles they played on each team.

Qualification signals. Although recruiters can assess a candidate's knowledge and experience by simply reading a set of skills and expertise, the information on professional networking sites adds valuable information for potential employers. Whereas traditional CVs might only mention the most recent positions, a professional social networking site can list an individual's entire work history. In addition, recruiters can evaluate the identity, skills, or trustworthiness of those giving recommendations by contacting them directly.

Reputation signals. LinkedIn recommendations are similar to traditional references, stating the relation between the recommender and the applicant and the applicant's position at the time they worked together. We can consider these

FOCUS: SOCIAL NETWORKING

as equivalent to direct references because they don't require additional interpretation steps.

Recommendations. Professional networking sites represent an excellent source of value to recruiters for both qualifications and reputations. However, candidates with little or no official experience with employers or accredited educational institutions might not be able to credibly list skills or qualifications, which limits their visibility. Other candidates might refrain from participation in such sites because they are successful enough without them. Recruiters should be aware of this bias and therefore embrace the full spectrum of social media sites.

Recruiters and candidates should also be aware of the leniency bias of online recommendations. Some software developers have become weary of such courtesy recommendations on LinkedIn, lowering the recommendations' value. In addition, very strong comments against a candidate are unlikely because they are publicly visible, including to the candidate personally.

Profile Aggregators

Profile aggregators create condensed profiles of individuals by collecting data from other sites. Masterbranch and Coderwall are examples of aggregators for developers. A more general aggregator, Yatedo, automatically creates a profile for any person it can find on the Web. Even though Masterbranch supports adding private source code to a member's profile, developer profile aggregators support mostly open source projects. This might discourage developers predominantly active in closed source projects from joining, potentially creating a significant bias in the available profiles.

Qualification signals. With aggregators, the most popular developers will have

the most visible profiles. This reduces the effectiveness of less vocal or active developers' profiles. Nonetheless, all developers have the potential to increase their visibility by proactively populating their online presence and having their potentially diverse experiences summarized. Aggregated profiles provide insight into development efforts distributed across multiple open source repositories and relate the knowledge they have to developers' Q&A profiles or their professional experience.

Reputation signals. Many aggregators support their own profile-building mechanisms. For instance, Coderwall awards achievement badges to developers on the basis of the code they commit to open source repositories. In addition, developers can endorse peers. However, even among developers, there's disagreement and uncertainty regarding such endorsements' value. Under some conditions, developers feel flattered to be endorsed. Others discard endorsements as meaningless because of the explicit prompt Coderwall gives to endorse others. One interviewee said, "I think there's some value to this kind of thing [Coderwall], but I'm not sure where it is yet."8

Recommendations. Profile aggregators' ability to collect and relate data about candidates makes them useful sources for information about developers' qualifications, especially as a starting point for deeper investigations. Similar to social code-sharing sites, nontechnical recruiters might find it difficult to interpret all the elements used on aggregated profiles and might benefit from the judgments of another developer.

Social Media as a Whole

The activity and experience documented in social media are easily verifiable, skills can be more readily assessed, and qualifications can in some cases be gathered with more confidence than asking a candidate to provide a set of references and list of qualifications.

Recruiters

We strongly recommend nontechnical recruiters to become familiar with social media, especially with sites that developers use. (The sites we discuss in this article are a good place to start.) When recruiters think critically about the positive and negative signals found in social media, they will be able to more easily assess and access developers. Moreover, this initiative should improve the image of recruiters among developers (a lack of familiarity with the developer communities has been shown in at least one case to create a negative perception of recruiters among developers8).

Although we believe that profile aggregators are especially useful for recruiters, they aren't yet comprehensive. Consequently, recruiters should use them only as a starting point to more thoroughly investigating the social media sites that provided the data in the candidates' profiles.

Candidates

The main challenge for candidates without official degrees or with limited field experience is to be recognized by recruiters. Social media have much to offer in this regard. Good work by individuals who made it visible to the public has and will be considered by at least some recruiters.² Developers with higher qualifications can also profit from participating in social media and public software development—for example, this can provide recruiters with insights about soft skills.⁸

evelopers who publish their work, even personal projects, document their progress and demonstrate that they're able to learn

new technologies. They can also learn best practices and try out diverse technologies and approaches to problems. We advocate the analysis of social media activity as one form of assessing qualifications and checking references for job candidates. However, recruiters should be wary of the inherent flaws associated with the use of social media. such as distorted or misrepresented reputation levels-for instance, when candidates overestimate their participation in large, well-known projects.

Social media are still in their infancy, and recruiters as well as job seekers are only starting to use them consistently for employment purposes. Once this field starts maturing, it will need to overcome many challenges—for example, how to filter out the noise of excessive, bloated online reputations, or how to compare similar candidates with different indicators of online experiences. Also, how recruiters accept these indicators and how to formalize a unique, accepted template to summarize online experience is a sociotechnical challenge that must still be addressed.

References

- 1. N.G. Barnes and E. Mattson, "Social Media in the 2009 Inc. 500: New Tools & New Trends," 2009; www.sncr.org/sites/default/ files/socialmedia2009_0.pdf.
- 2. T.J. Hayes, D. Ruschman, and M.M. Walker, "Social Networking as an Admission Tool: A Case Study in Success," J. Marketing for Higher Education, vol. 19, no. 2, 2009, pp. 109 - 124.
- 3. D. Joseph et al., "The Career Paths Less (or More) Traveled: A Sequence Analysis of IT Career Histories, Mobility Patterns, and Career Success," MIS Q., vol. 36, no. 2, 2012, pp. 427-452.
- 4. M.-A. Storey et al., "The Impact of Social Media on Software Engineering Practices and Tools," Proc. FSE/SDP Workshop Future of Software Engineering Research (FoSER 10), ACM, 2010, pp. 359-364.
- 5. D.M. Boyd and N. Ellison, "Social Network Sites: Definition, History, and Scholarship," J. Computer-Mediated Comm., vol. 13, no. 1, 2007, pp. 210-230.
- 6. H.C. Stuart et al., "Social Transparency in Networked Information Exchange: A Theoret-



ANDREA CAPILUPPI is a lecturer in the School of Information Systems, Computing and Mathematics at Brunel University. His research interests include software maintenance and evolution, software architectures, component reuse, and open source software. Capiluppi received a PhD in computer systems from Politecnico di Torino. Contact him at andrea.capiluppi@brunel.ac.uk or on Twitter @acapiluppi.



ALEXANDER SEREBRENIK is an assistant professor of modeldriven software engineering at Eindhoven University of Technology. His research interests include software evolution and maintenance, social media, program analysis, and transformation. Serebrenik received a PhD in computer science from Katholieke Universiteit Leuven. He's a member of IEEE and the European Research Consortium for Informatics and Mathematics Working Group on Software Evolution. Contact him at a.serebrenik@tue.nl or on Twitter @aserebrenik.



LEIF SINGER is a PhD student in software engineering at Leibniz Universität Hannover. His research interests include the systematic use of social software to support software engineers, such as using newsfeeds, game mechanics, or reputation systems. Singer received an MSc in computer science from Leibniz Universität Hannover. Contact him at leif.singer@inf.uni-hannover.de or on Twitter @lsinger.

- ical Framework," Proc. ACM Conf. Computer Supported Cooperative Work (CSCW 12), ACM, 2012, pp. 451-460.
- 7. L. Dabbish et al., "Social Coding in GitHub: Transparency and Collaboration in an Open Software Repository," Proc. ACM Conf. Computer Supported Cooperative Work (CSCW 12), ACM, 2012, pp. 1277-1286.
- 8. L. Singer et al., Mutual Assessment in the Social Programmer Ecosystem: An Empirical Investigation of Developer Profile Aggregators, tech. report DCS-347-IR, Univ. Victoria, Canada, 2012.
- 9. T. Schmader, J. Whitehead, and V.H. Wysocki, "A Linguistic Comparison of Letters of Recommendation for Male and Female Chemistry and Biochemistry Job Applicants," Sex Roles, vol. 57, nos. 7-8, 2007, pp. 509-514.
- 10. J.M. McCarthy and R.D. Goffin, "Improving the Validity of Letters of Recommendation: An Investigation of Three Standardized Reference Forms," Military Psychology, vol. 13, no. 4, 2001, pp. 199-222.
- 11. J. Antin and E. Churchill, "Badges in Social Media: A Social Psychological Perspective," CHI 2011 Gamification Workshop, ACM, 2011, pp. 1-4.
- 12. K. Stein and C. Hess, "Does It Matter Who Contributes: A Study on Featured Articles in

- the German Wikipedia," Proc. 18th Conf. Hypertext and Hypermedia (HT 07), ACM, 2007, pp. 171-174.
- 13. C. Steinfield, N. Ellison, and C. Lampe, "Social Capital, Self-Esteem, and Use of Online Social Network Sites: A Longitudinal Analysis," J. Applied Developmental Psychology, vol. 29, no. 6, 2008, pp. 434-445.
- 14. L. Mamykina et al., "Design Lessons from the Fastest Q&A Site in the West," Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI 11), ACM, 2011, pp. 2857-2866.
- 15. S.C. Doty-Navarro and B.H. Kleiner, "How to Effectively Check References and Perform Background Investigations of Job Applicants," Management Research News, vol. 23, nos. 7–8, 2000, pp. 56–62.

