
Recruiters, Job Seekers and Spammers: Innovations in Job Search at LinkedIn

Daria Sorokina
dsorokina@linkedin.com

LinkedIn

Keynote Abstract

In this talk I will share insights about how LinkedIn responds to the needs of two groups of users - recruiters and job seekers, while making life more difficult for a third group - search engine optimization (SEO) spammers.

When recruiters search for good candidates, they are primarily interested in candidates whose profiles match the descriptions of a particular job posting. However, a simple textual match with the job requirements is often not enough. We will discuss other signals that can help recruiters to achieve their goals along with the ways we combine the signals from multiple models.

Professional search at LinkedIn is highly personalized. Regardless of what they enter into the search box, job seekers prefer to find job postings whose requirements match their skills and generally are more interested in opportunities near where they currently live. This personalization makes standard crowdsourced labeling of the training data complicated - it is hard for a random labeler to imagine which job ad a particular user would find relevant. As a result, our job search models have to rely primarily on signals from user activity.

Being ranked high in the results for common queries on LinkedIn search can bring new clients, business connections, or job offers. As a result, some users create profiles stuffed with keywords, fake entries and other junk information in order to raise their search ranking. While we can not share specific details on how we identify such users, I will present the general machine learning framework along with examples of difficult cases and unusual spammer techniques we have encountered.

About the speaker

Daria Sorokina is Senior Data Scientist at LinkedIn, who enjoys working with large complex data sets. For the last few years she has been specializing in search, before that she worked on a variety of machine learning applications in different domains. Daria is the author of Additive Groves, the best off-the-shelf machine learning algorithm for a variety of tasks.

Copyright © by the paper's authors. Copying permitted only for private and academic purposes.

In: M. Lupu, M. Salampasis, N. Fuhr, A. Hanbury, B. Larsen, H. Strindberg (eds.): Proceedings of the Integrating IR technologies for Professional Search Workshop, Moscow, Russia, 24-March-2013, published at <http://ceur-ws.org>