

Structuring the Unstructured: From Startup to Making Sense of eBay's Huge eCommerce Inventory

Ido Guy and Kira Radinsky
eBay Research
Netanya, Israel
idoguy@acm.org,kradinsky@ebay.com

ABSTRACT

Electronic commerce continues to gain popularity in recent years. On eBay, one of the largest on-line marketplaces in the world, millions of new listings (items) are submitted by a variety of sellers every day. This renders a rich diverse inventory characterized by a particularly long tail [7]. In addition, many items in the inventory lack basic structured information, such as product identifiers, brand, category, and other properties, due to sellers' tendency to input unstructured information only, namely title and description [6]. Such inventory therefore requires a handful of large-scale solutions to assist in organizing the data and gaining business insights.

In 2016, eBay acquired SalesPredict to help structure its unstructured data. In this proposed presentation, we will share the story of a research startup from its inception until its acquisition and integration as eBay's data science team. We will review the numerous challenges from research and engineering perspectives of a startup and the principal challenges the eBay data science organization deals with today. These include the identification of duplicate, similar, and related products; the extraction of name-value attributes from item titles and descriptions; the matching of items entered by sellers to catalog products; the ranking of item titles based on their likelihood to serve as "good" product titles; and the creation of "browse node" pages to address complex search queries from potential buyers. We will describe how the eBay data science team approaches these challenges and some of the solutions already launched to production. These solutions involve the use of large-scale machine learning, information retrieval, and natural language processing techniques, and should therefore be of interest to the SIGIR audience at large.

1 ABOUT THE PRESENTERS

Ido Guy is the head of EMEA research at eBay, leading teams in Israel and Germany whose expertise lie in text analysis. Prior to joining eBay, Ido spent three years at Yahoo Research as a Principal Research Engineer and before that he managed the Social Technologies group at IBM Research. In all of these roles, the focus of his work was building state-of-the-art systems that deliver innovation into products. Ido has published over 50 papers in leading conferences and journals, spanning human-computer interaction [4], recommender systems [3], and information retrieval [2, 5]. His

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SIGIR'17, August 7–11, 2017, Shinjuku, Tokyo, Japan

© 2017 Copyright held by the owner/author(s). 978-1-4503-5022-8/17/08.

DOI: 10.1145/nnnnnnn.nnnnnnn

SIGIR 2016 study of voice search received the best paper honorable mention award [1].

Kira Radinsky is the chief scientist and director of data science of eBay, where she is building the next generation predictive data mining, deep learning, and natural language processing solutions for eCommerce. Dr. Radinsky co-founded SalesPredict, acquired by eBay in 2016. She gained international recognition for her work at Microsoft Research, where she developed predictive algorithms that recognized the early warning signs of global events, including political riots and disease epidemics. In 2013, she was named to the MIT Technology Review 35 young innovators under 35, and in 2015 as Forbes 30 under 30 rising stars in enterprise technology. She is a frequent presenter at global tech events, including TEDx and the Harvard Business Review. Radinsky also serves as a visiting professor at the Technion, Israel's Institute of Technology, where she focuses on the application of predictive data mining in medicine.

2 ABOUT THE COMPANY

SalesPredict was founded in 2012 as a pioneer in the field of predictive marketing. It created a technology that harvests the unstructured Web and gathers information about interactions between companies and individuals. Combining it with internal sales data of its customers, SalesPredict created one of the largest economic graphs providing predictive analysis of future business interactions. On average, conversion rates of its customers went up by 300%. In 2016, it was acquired by eBay.

eBay is one of the world's largest on-line marketplaces, allowing hundreds of millions of sellers and buyers to exchange goods of virtually any kind. Data science on top of eBay's inventory aims to arrange, enrich, and gain insights from the large data in order to enhance the shopping experience and support services such as faceted navigation, search, and recommendation.

REFERENCES

- [1] Ido Guy. 2016. Searching by Talking: Analysis of Voice Queries on Mobile Web Search. In *Proc. SIGIR*. 35–44.
- [2] Ido Guy, Avihai Mejer, Alexander Nus, and Fiana Raiber. 2017. Extracting and Ranking Travel Tips from User-Generated Reviews. In *Proc. WWW*. 987–996.
- [3] Ido Guy, Inbal Ronen, Elad Kravi, and Maya Barnea. 2016. Increasing Activity in Enterprise Online Communities Using Content Recommendation. *ACM Trans. Comput.-Hum. Interact.* 23, 4, Article 22 (2016), 22:1–22:28 pages.
- [4] Ido Guy, Inbal Ronen, Naama Zwerdling, Irena Zuyev-Grabovitch, and Michal Jacovi. 2016. What is Your Organization 'Like'? A Study of Liking Activity in the Enterprise. In *Proc. CHI*. 3025–3037.
- [5] Elad Kravi et al. 2016. One Query, Many Clicks: Analysis of Queries with Multiple Clicks by the Same User. In *Proc. CIKM*. 1423–1432.
- [6] Karin Mauge, Khash Rohanimanesh, and Jean-David Ruvini. 2012. Structuring e-Commerce Inventory. In *Proc. ACL*. 805–814.
- [7] Neel Sundaresan. 2011. Recommender Systems at the Long Tail. In *Proc. RecSys*. 1–6.