Constructing and Mining Web-Scale Knowledge Graphs

KDD 2014 Tutorial

Antoine Bordes
Facebook
770 Broadway
New York, NY 10003
abordes@fb.com

Evgeniy Gabrilovich
Google
1600 Amphitheatre Parkway
Mountain View, CA 94043
gabr@google.com

ABSTRACT

Recent years have witnessed a proliferation of large-scale knowledge graphs, such as Freebase, YAGO, Google's Knowledge Graph, and Microsoft's Satori. Whereas there is a large body of research on mining homogeneous graphs, this new generation of information networks are highly heterogeneous, with thousands of entity and relation types and billions of instances of vertices and edges. In this tutorial, we will present the state of the art in constructing, mining, and growing knowledge graphs. The purpose of the tutorial is to equip newcomers to this exciting field with an understanding of the basic concepts, tools and methodologies, available datasets, and open research challenges. A publicly available knowledge base (Freebase) will be used throughout the tutorial to exemplify the different techniques.

Who should attend

Over the recent years, multiple large-scale knowledge bases have come to existence, from purely academic projects such as YAGO to major commercial projects such as Facebook's Open Graph, Google's Knowledge Graph, and Microsoft's Satori. These knowledge bases are an enabling resource for a plethora of new knowledge-rich applications. The tutorial is designed to provide an overview of such applications and use-cases, as well as equip newcomers to this area with solid understanding of the basic techniques, research challenges, and pointers to relevant literature. This tutorial will bridge together three top-level areas of the KDD 2014 Call for Papers, namely, algorithms, applications, and big data.

Prerequisites

The tutorial will be self-contained, and is designed to be accessible to researchers and practitioners in related disciplines who want to learn more about mining large-scale knowledge graphs. An undergraduate degree in CS or statistics should be sufficient to understand most of the content in the tutorial.

KDD'14, August 24–27, 2014, New York, NY, USA.
ACM 978-1-4503-2956-9/14/08.
http://dx.doi.org/10.1145/2623330.2630803.

Instructors

Antoine Bordes is a staff research scientist at Facebook Artificial Intelligence Research. Prior to joining Facebook in 2014, he was a CNRS research staff in the Heudiasyc laboratory of the University of Technology of Compiègne in France. In 2010, he was a postdoctoral fellow in Yoshua Bengio's lab at Universite de Montreal. He received his PhD in machine learning from Pierre & Marie Curie University in Paris in early 2010. From 2004 to 2009, he collaborated regularly with the Machine Learning department of NEC Labs of America in Princeton. He received two awards for best PhD from the French Association for Artificial Intelligence and from the French Armament Agency as well as a Scientific Excellence Scholarship by CNRS in 2013. Antoine is a pioneer in the use of embedding models for modeling knowledge bases and co-authored many papers on the topic in recent years. He is also a specialist of natural language processing, deep learning and large-scale learning.

Evgeniy Gabrilovich is a senior staff research scientist at Google, where he works on knowledge discovery from the web. Prior to joining Google in 2012, he was a director of research and head of the natural language processing and information retrieval group at Yahoo! Research. Evgeniy is an ACM Distinguished Scientist (2012), and is a recipient of the 2010 Karen Sparck Jones Award for his contributions to natural language processing and information retrieval. He is also a recipient of the 2014 IJCAI-JAIR Best Paper Prize. Evgeniy serves as a PC chair for WSDM 2015, and has served as an area chair or senior program committee member at numerous major conferences, including SIGIR, KDD, WWW, WSDM, AAAI, IJCAI, ACL, EMNLP, CIKM, ICDM and ICWSM. He has organized a number of workshops and taught multiple tutorials at SIGIR, ACL, WWW, WSDM, ICML, IJCAI, AAAI, CIKM, and EC. Evgeniy earned his PhD in computer science from the Technion - Israel Institute of Technology.