Active Objects: An Entity-Centric Search Experience

Speaker

Patrick Pantel 氏 (Microsoft Research)

Patrick Pantel is a Senior Researcher at MSR, conducting research in large-scale natural language processing, text mining, web search, and knowledge acquisition. Prior he served as a Senior Research Manager at Yahoo! Labs, and as a Research Assistant Professor at the USC Information Sciences Institute. In 2003, he received a Ph.D. in Computing Science from the University of Alberta in Edmonton, Canada.

Abstract

We introduce an entity-centric search experience, called *active objects*, in which entity-bearing information requests are paired with actions that can be performed on the entities, thus enabling the brokering of web pages and applications on the Web that can satisfy the intended action.

In this vision, the broker is aware of all entities and actions of interest to its users, understands the intent of the user, ranks all providers of actions, and provides direct actionable results through APIs with the providers. For example, consider a user who queries for "jetbeam rrt-0," a flashlight. The broker would recognize the particular entity mentioned in the query, would return a personalized ranked list of actions to the user, would allow the user to save clicks and time to accomplish his intended action, and even would allow the user to sometimes discover new actions to help them toward their goals. New revenue streams open up from paid action placement, lead generation, and on-site commercial transactions.

In an annotation study conducted over a random sample of user query sessions, we found that a large proportion of queries in query logs involve actions on entities, calling for an automatic approach to identifying relevant actions for entity-bearing queries. We pose the problem of finding actions that can be performed on entities as the problem of doing probabilistic inference in a graphical model that captures how an entity bearing query is generated. We design models of increasing complexity that capture latent factors such as entity type and intended actions that determine how a user writes a query in a search box, and the URL that they click on. Given a large collection of real-world queries and clicks from a commercial search engine, the models are learned efficiently through maximum likelihood estimation using an EM algorithm. Given a new query, probabilistic inference enables recommendation of a set of pertinent actions and hosts. We propose an evaluation methodology for measuring the relevance of our recommended actions, and show empirical evidence of the quality and the diversity of the discovered actions. We further show how the models can be used to enhance entity repositories such as Freebase with type distributions over an application domain such as web search. We end by showing how the proposed models can be cast in a wider framework providing a generalized entity-centric natural language user interface, where experiences with entities can be rendered consistent across information devices.