

Specialized Query Understanding

Donald Metzler
Information Sciences Institute
University of Southern California

July 23, 2010
SIGIR 2010 Query Representation and Understanding Workshop



Importance of Query Understanding

- Query understanding is important for driving search technologies to the next level
- So important that commercial search companies have made up various catchphrases about it:
 - “Decision engine” (Bing)
 - “The end of the 10 blue links” (Yahoo!)
 - “Best locally relevant results served globally” (Google)
- These are just different ways of saying their goal is to “understand what users want and give it to them”

Query Understanding Perspectives

- IR community
 - Query representations
 - Similarity measures / features
- NLP community
 - Knowledge representation
 - Semantics
- Web search community
 - Query logs
 - User behavior
- Machine learning community
 - Models
 - Optimization procedures

Search Research Flavors

- Two general flavors of search research
- “One size fits all” approaches
 - Focus on a general task or class of tasks
 - Makes a minimal number of assumptions
 - Effectiveness likely to have high variance across tasks
- Specialized approaches
 - Focus on a specific task or aspect of a task
 - Exploits characteristics of task or domain
 - Effectiveness often improved over “one size fits all” approaches

Specialized Query Understanding

- Effective search requires “general specialization”
- “One size fits all” approaches are too generic
- Specialized approaches are too specific
- Need to integrate techniques from multiple research perspectives
- Develop modular query understanding frameworks

General Specialization

- There are many different ‘classes’ of queries
- Each class varies across the following dimensions:
 - Query representations
 - Similarity measures / features
 - Ranking functions
 - Training data
 - Evaluation metrics

Manual Specialization

- Specialization can be achieved by many small teams of search researchers, scientists, and engineers
- Responsibility of each group
 - Identify interesting query class
 - Feature engineering / model development
 - Devise appropriate evaluation metrics
 - Integrate with rest of the search system

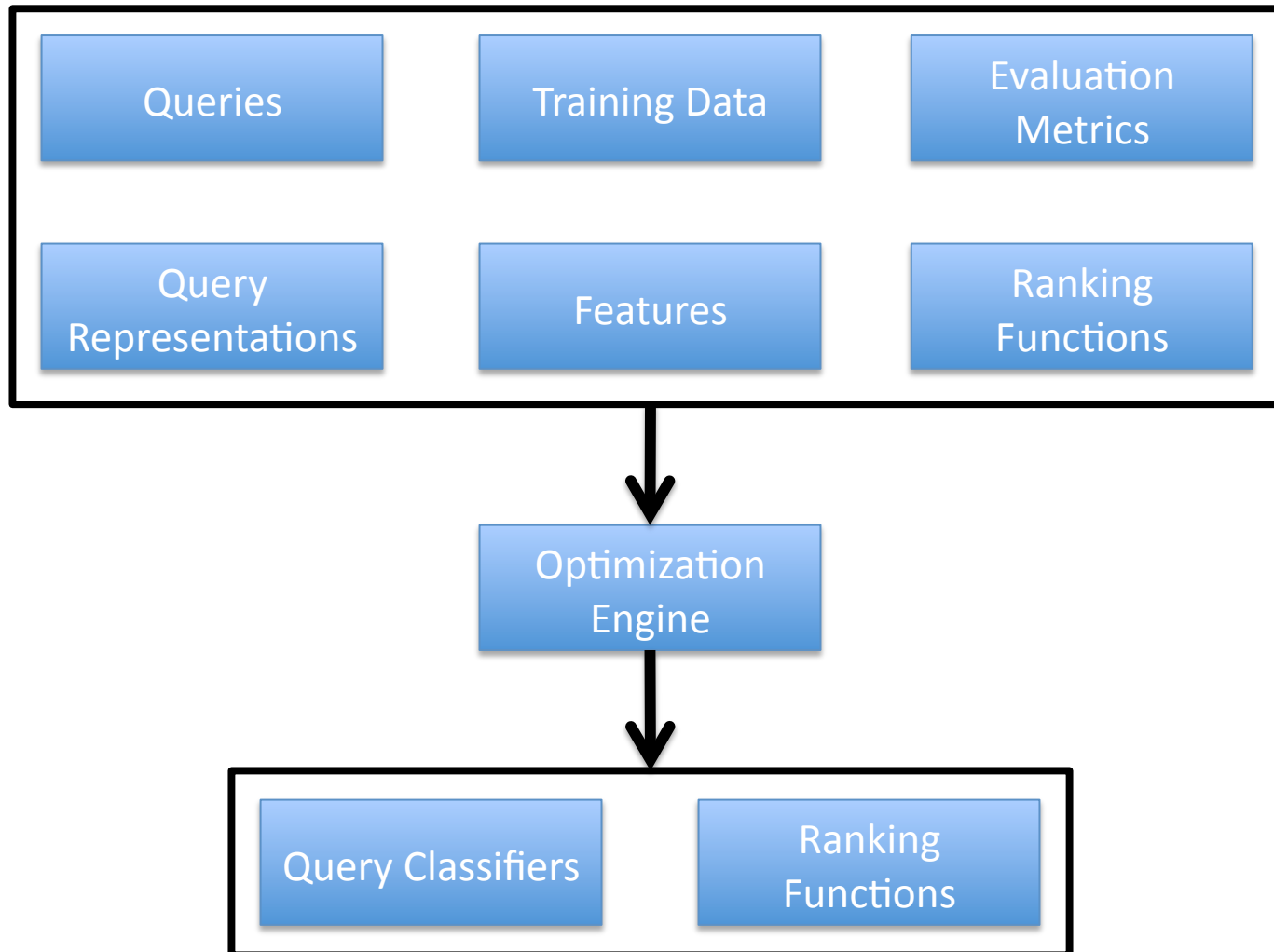
Examples of Manual Specialization

- News search queries
- Factoid questions
- Navigational queries
- Long queries
- Temporal queries
- ...

Automatic Specialization

- Manual specialization is costly, error-prone, and does not easily scale to many query classes
- Many benefits to (semi-)automatic deep specialization
- Challenging problem with many interesting research directions for both academia and industry

Automatic Specialization



Research Challenges

- Expressive query representations
- Advanced feature engineering
- Ranking function construction
- Generic classes of evaluation metrics
- Novel sources of relevance information

Research Challenges

- Automatically determining best evaluation metric for each class of queries
- Flexible query processing strategies that can be adapted to specialized queries
- Robust, modular framework for integrating all of these pieces
- Machine learning methods for optimizing end-to-end system

Questions?