

Mineração na Web: Roteamento de Consultas

Daker Fernandes
Guilherme Cavalcanti
{dfp,goc2}@cin.ufpe.br

Introdução

- O que é roteamento de consultas?



Introdução

- O que é roteamento de consultas?



Introdução

- O que é roteamento de consultas?



Histórico

- Bancos de Dados Distribuídos
 - Roteamento de Consultas por Tabela de Roteamento

- PDMS - P2P Data Management System
 - Roteamento de Consultas por Similaridades de Esquema

Onde entra a mineração na Web?

- P2P-RI
- Web Search
- Social Query

P2P-RI

- Sistema de RI com a arquitetura P2P
- Escalabilidade
- Tolerância a falhas
- Roteamento de Consultas por localização de Recursos
 - DHT
 - Similaridade de esquemas

Web Search

- Search engines são melhores quando trabalham com tópicos específicas
- "Qual o melhor engenho de busca para a minha consulta?"
- Solução: Meta-Busca
- Interface Única

Web Search Processo

1. Pré-processamento da consulta
 1. Remoção de palavras irrelevantes na consulta
 2. Normalização de Termos
 1. Sem normalização
 2. Tesouro
 3. Semântica Latente

Web Search Processo

2. Expansão da Consulta
 1. Adição de Termos para enriquecimento da consulta

3. Extração de Tópicos da Consulta
 1. Índice de Termos
 2. Clustering de Termos
 3. Técnicas de Classificação por Aprendizagem
 1. SVM!

Web Search Processo

4. Recuperação de Search Engines

1. Índice Invertido
2. Classificador

5. Reformulação de Consulta

1. Conversão dos Termos da Consulta
2. Roteamento da Consulta para as Outras Engines

6. Unificação de Resultados

1. Normalização dos resultados em uma interface única

Web Search Quem usa?

- Search.com
- Google!
- Q-Pilot



Social Query

- A consulta pode ser muito específica
 - Alternativas
 - Roteamento para *search engines* (Q-Pilot)
 - Roteamento para usuários com *expertise* para tal tópico

"Do you have any good babysitter recommendations in Palo Alto for my 6-year-old twins? I'm looking for somebody that won't let them watch TV."

Social Query *The library and the village*

- O paradigma da biblioteca
 - Consulta através de keywords
 - Confiança baseada na autoridade
- O paradigma da vila
 - Consulta através de linguagem natural
 - Confiança baseada na intimidade

... real-time responses from socially proximal responders tend to elicit highly contextualized and subjective queries.

The Anatomy of a Large-Scale Social Search Engine

Social Query Processo

1. Pré-processamento
 - Inicialização do usuário (*signup time*)
2. Análise da consulta
 - É mesmo uma pergunta?
 - Extração de tópicos
3. Ranking de usuários
 - Composição entre
 - *Score* de relevância (*query-dependent*)
 - *Score* de qualidade (*query-independent*)
4. Roteamento da pergunta

Social Query Análise da consulta

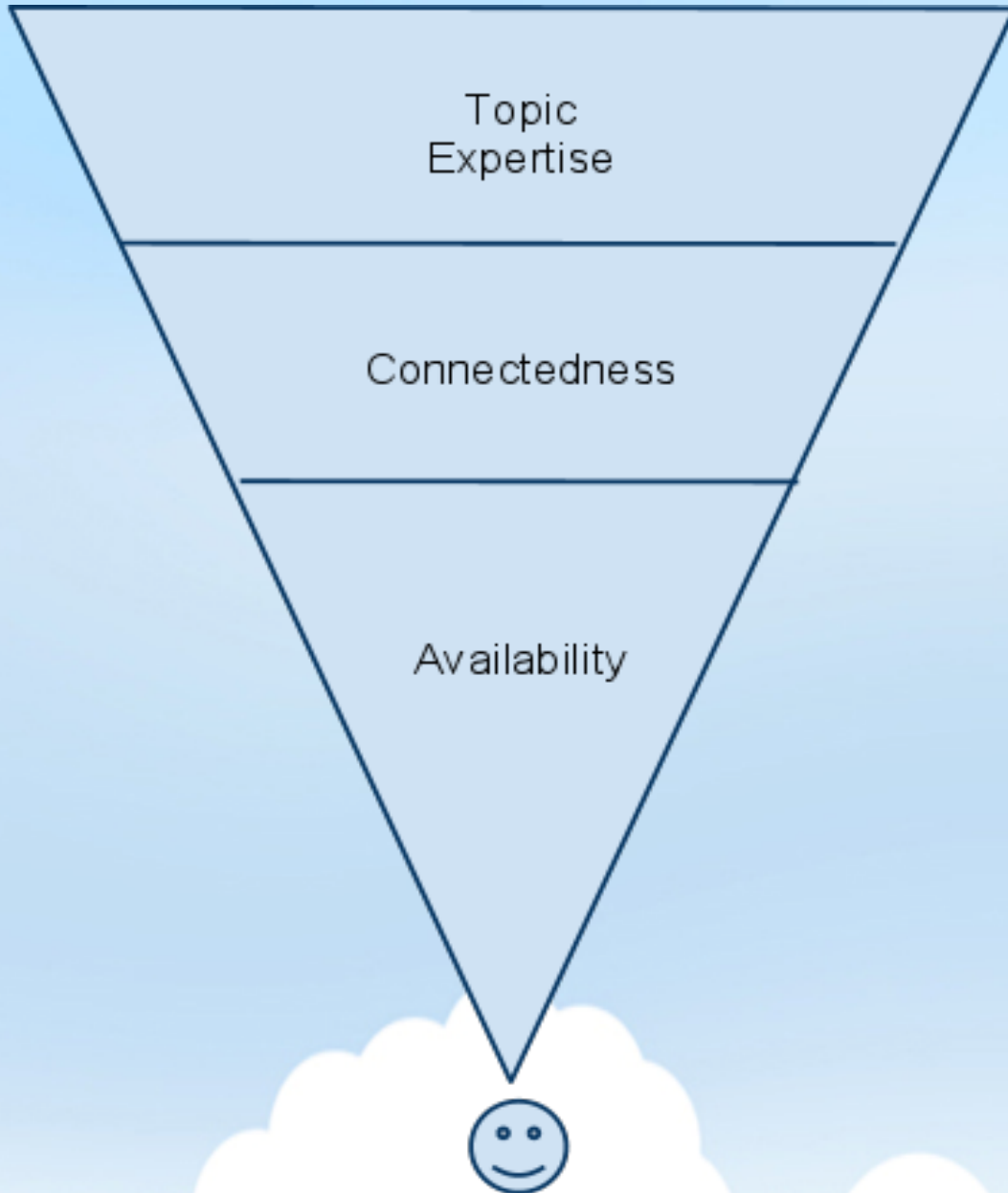
- Filtragem
 - Questão inapropriada
 - Questão trivial
 - *Location Sensitive*
- Ranking de **tópicos** específicos para uma dada questão
 - Pode ser uma combinação dos resultados de vários classificadores
 - Calculado em tempo de execução

*The purpose of the Question Analyzer is to determine a **scored list** of topics $p(t|q)$ for each question q representing the **semantic subject matter of the question**.*

Social Query *Ranking* de usuários

- **Tópicos** de expertise do **usuário**
 - Informado pelo usuário e seus amigos
 - Análise de perfil em outras redes sociais
- **Conectividade** do usuário
 - Ponto de vista social
 - Ponto de vista demográfico
 - Outras características mais subjetivas
 - Educação, *verbosity*, similaridade de perfil

Social Query Roteamento

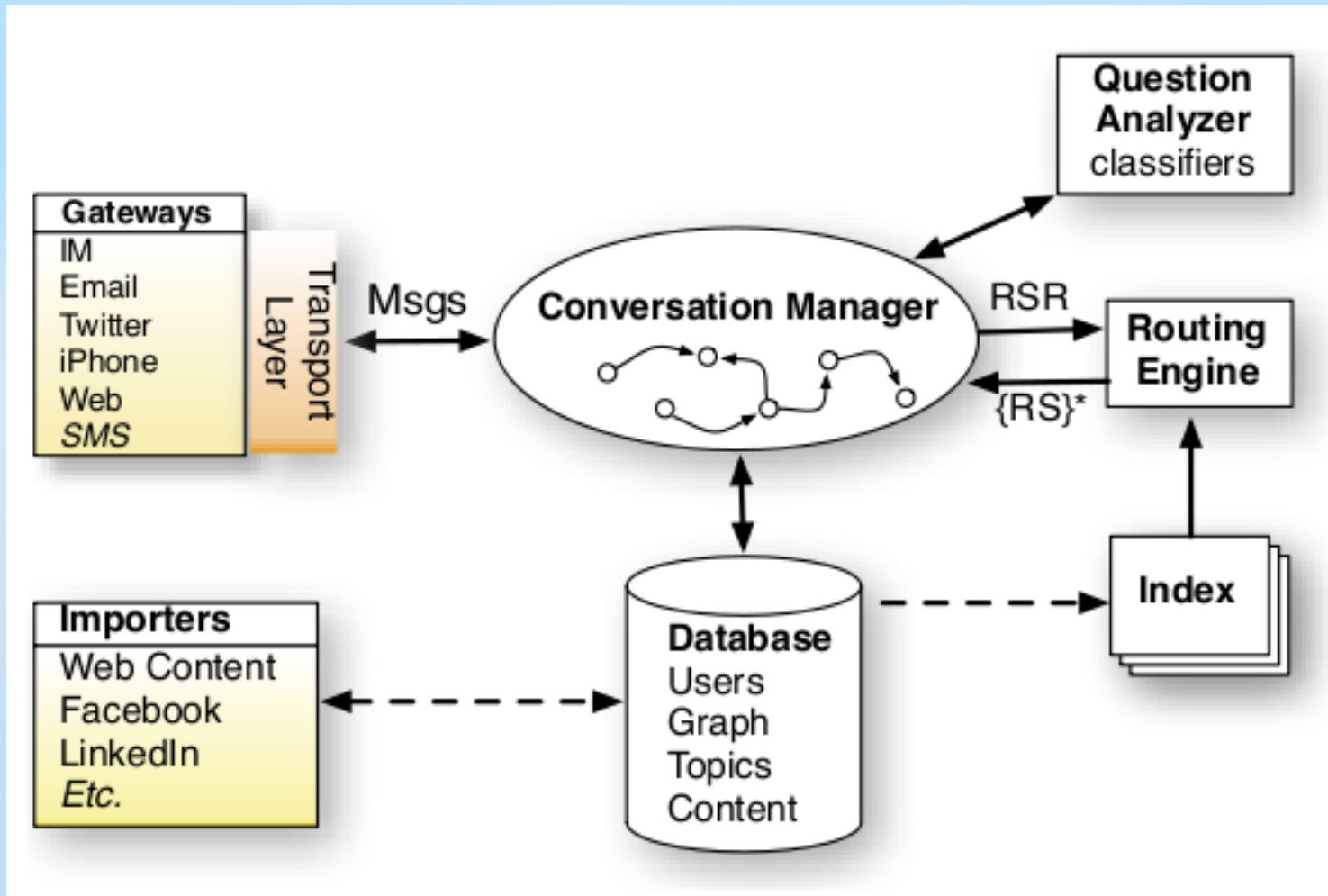


Subconjunto de usuários que possuem expertise no tópico da pergunta.

Subconjunto mais "próximos" de quem pergunta

Subconjunto dos usuários com disposição para responder perguntas num dado momento

Social Query Processo

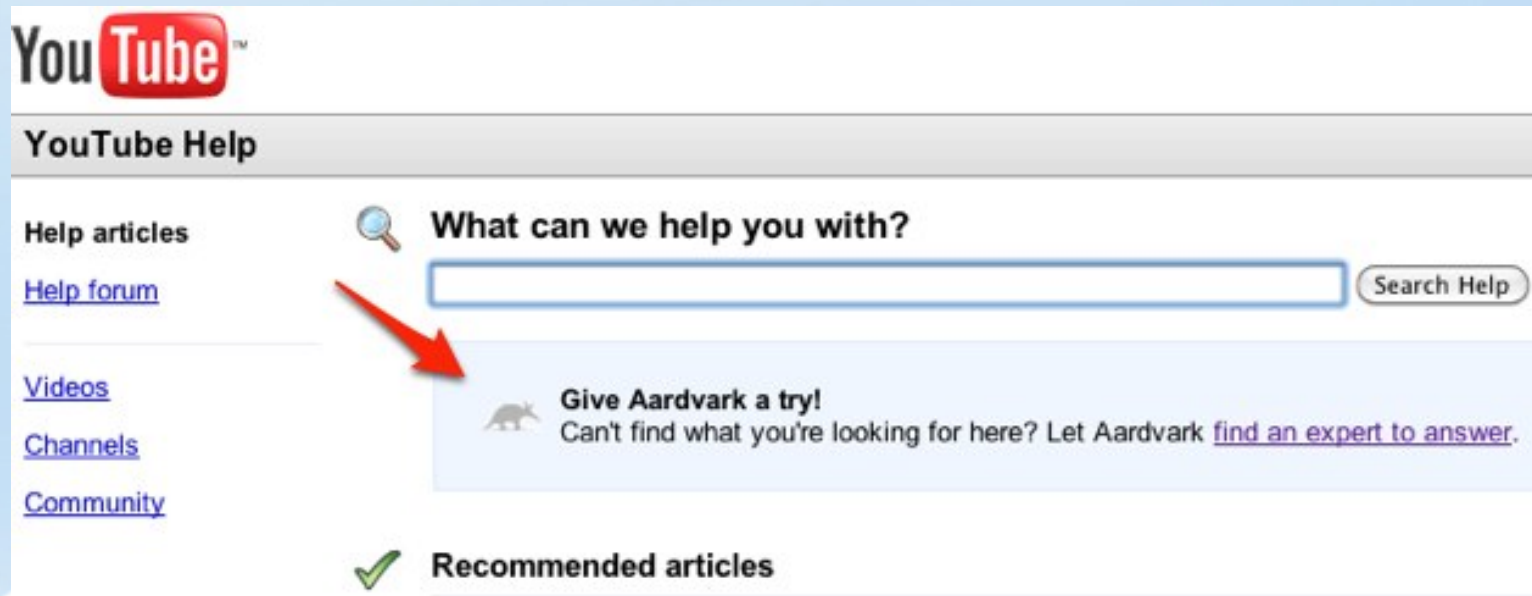


Social Query: Quem usa?

- Aardvark (vark.com)
- *Help* do Google

Google Puts Its \$50 Million To Work, Starts Using Aardvark For Help Support

[TechCrunch](#)



Conclusão

- **Search engines** de propósito geral sempre retornam documentos. Porém não necessariamente suprem a necessidade de informação dos usuários.
- **Social search engines** nem sempre retornam respostas, porém, na maioria dos casos, suprem a necessidade de informação dos usuários.
 - Tende a ter bons resultados em contextos específicos
 - Pode ser melhorada através do estímulo aos usuários

Referências

1. Query Routing for Web Search Engines: Architecture and Experiments. Atsushi Sugiura and Oren Etzioni
2. The Anatomy of a Large-Scale Social Search Engine. Damon Horowitz and Sepandar D. Kamvar
3. REMINDIN': Semantic Query Routing in Peer-to-Peer Networks Based on Social Metaphors. Christoph Tempich, Steffen Staab and Adrian Wranik
4. A Social Query Model for Decentralized Search, Arindam Banerjee and Sugato Basu
5. An Architecture for Peer-to-Peer Information Retrieval
Karl Aberer, Fabius Klemm, Martin Rajman, Jie Wu