Big Search: the key to unlock the true value of Big Data

Xi Zhang
Key Laboratory of Trustworthy Distributed Computing and Service,
Ministry of Education
Beijing University of Posts and Telecommunications
CAFOE 2015, Irvine, California
Outline

1. Motivation
2. Concept and New Features
3. Key Issues
4. Conclusion
Motivation

Big data is everywhere: Industries are digging big data for hidden values.

Personalized recommendation system brings 30% extra sales to Amazon

TARGET recommends products to customers by mining purchase history records, which increases the income by about $23 billions.
However, digging is mostly domain-specific, and transferring across domains is rather difficult.

**Question:** Can we achieve more by integrating various big data from multiple domains?

- **Risk Control by Mining**
  - Purchase history & Property

- **Etiology Discovering by Mining**
  - Dietary & Medical Records
Third-parties are required to be responsible for providing the cross-domain service.

Who is best suitable to take the role?
Search engine, but existing ones may not meet the requirements in big data era.

Big data era requires **Big Search** engine.
The extension of search space:
From solo internet to ubiquitous cyberspace which connects people, information and objects.
The evolution of web applications:
From Web1.0 to social media applications of Web2.0, and towards the Intelligent Web3.0.
Challenges of Big Search Engine

The abundance of data:
Transition from static web pages to dynamic, fast generated multimodal data.
5Vs (Volume, Velocity, Variety, Veracity and Value)

- **Static webpage**
- **IoT data**
- **SN data**
- **Mobile internet data**
- **A/V data**
- **S/T data**
- **Surveillance Camera**
More and more user requirements:
From keyword search to semantic oriented search, and towards intelligent solutions.
Big Search refers to a new intelligent search system for cyberspace, which can deal with various kinds of big data from web1.0 to web3.0, involving people, info. and things, aiming to find intelligent solutions.
New Features

1) Sourcing
   Sourcing from Cyber

2) Sensing
   Sensing in Context

3) Synthesizing
   Synthesizing Multi-channel

4) Security
   Privacy & Trust

5) Solution
   Intelligent Solution

Big Search

5S
Feature: **Sourcing from Cyber**

- **Sourcing**
- **Sensing**
- **Synthesizing**
- **Security**
- **Solution**

➢ to get data from the Cyberspace

**Traditional Search**

- just crawls data from the Web/Internet

**VS**

**Big Search**

- gets a variety of information from multiple networks involving people, objects, events etc
Feature: Sensing in Context

- to accurately understand user intentions

Traditional Search mainly cares about the keywords, involving simple semantic analysis

Big Search understands user better by time, location, preferences, emotion or even gestures.

Semantic, Time, Location, Preference, Emotion
Feature: Synthesizing the Knowledge

- To integrate multi-modal information into accessible knowledge

Traditional Search

- gives the most relevant results based on a ranking algorithm

Big Search:

- builds knowledge warehouse and indexing from comprehensive entity-relationships of multi-modal data.
Feature: Security (Privacy & Trust)

- Credibility of search results, and privacy protection

Traditional Search

VS

Big Search

filters malicious and spam information

addresses more security issues: source credibility, secure access, privacy protection and precise filtering.
**Feature: Intelligent Solution**

- **Sourcing**
- **Sensing**
- **Synthesizing**
- **Secure**
- **Solution**

- **Smart Searching for answers, find smart answers for users**

  **Traditional Search**

  only provides existing information to meet user requirements

  **Big Search**

  finds smart solutions by integrating existing knowledge and inferring new knowledge

  **Travel program**

  **Medical program**
Big Search Engine: Bridging Information Gap

1. Information Acquisition
2. Intention Understanding
3. Knowledge Warehousing
4. Security and Trust
5. Knowledge Inference

Users
Ubiquitous CyberSpace

Information GAP

User intention

Search queries
Big Search Engine: 5 Key Issues

**Information Acquisition**
Collect, acquire various data efficiently and are able to discover relevant entities

**Intention Understanding**
Make the search engine understand users’ true demands

**Knowledge Warehouse**
Design efficient and scalable knowledge warehouse based on entity relationship model

**Security & Trust**
Provide trustworthy and privacy preserving services

**Knowledge Inference**
Inferring new knowledge based on existing knowledge and rules
Case study: What G & B Do

Google Search: where to see red leaves

About 392,000,000 results (0.51 seconds)

Images for where to see red leaves

More images for where to see red leaves

Autumn leaves in Japan: Where to see autumn leaves?
www.japan-guide.com/e/e2014_where.html Japan Guide
Aug 16, 2012 - Hiking is the most rewarding way to see the colorful leaves in the ... The trees around the temples large wooded stages usually turn red in the ...
Case study: What Big Search Do

Where to see red leaves?

Nov. 1th 6:30AM; BUPT, Beijing

<table>
<thead>
<tr>
<th>Places of Interests</th>
<th>Xiang Mountain</th>
<th>Mang Mountain</th>
<th>Badaling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Status</td>
<td>Green</td>
<td>Red</td>
<td>Fallen</td>
</tr>
<tr>
<td>Distance</td>
<td>20km</td>
<td>40km</td>
<td>120km</td>
</tr>
</tbody>
</table>

Information Acquisition: real time pictures and transportation
Intention Understanding: willing to see red leaves now
Knowledge Warehouse: places of interests relationship graph
Security & Trust: spam filtering & privacy preserving
Knowledge Inference: recommending the best place to visit
Outline

1. Motivation
2. Concept and New Features
3. Key Issues
4. Conclusion
Conclusion

Challenges
- From Info to People, Info, Things
- From Web 1.0 to Web2.0/3.0
- From Static data to Big data
- From Keyword to Intelligent Answer

New Features
- Sourcing
- Sensing
- Synthesizing
- Security
- Solution

Key Issues
- Information Acquisition
- Intention Understanding
- Knowledge Warehouse
- Privacy & Trust
- Knowledge Inference
Acknowledgement

All the contributors
Binxing Fang, Beijing University of Posts and Telecommunications
Manqing Wu, China Electronics Technology Group Corporation
Bin Zhou, Yan Jia, National University of Defense Technology
Lihua Yin, Institute of Information Engineering, CAS

Sponsor

Thanks
Contact: zhangx@bupt.edu.cn