

Continuous Topic Analysis

SYSTEMATIC THOUGHT LEADERSHIP FOR INNOVATIVE BUSINESS



Topic Tracker Team

Qi Li,

Supervisor: Mia Ming Mao,

Project Team: Topic Tracker @ Continuous Sense Making

Agenda

- Overview
- Statistical Approach
- Semantic Approach
- Demo
- Conclusion

Continuous Sense Making

- Continuous:
 - Information produces
 - Information analysis
- Sensemaking is the ability or attempt to make sense of an ambiguous situation. -- from wikipedia
- Assumption:
 - surplus computing power could be used to continuously analyze large multi-media datasets for relevant business information
 - surplus computing power will provide the performance to support analysis in Collaboration-based Information Systems.

http://research.pal.sap.corp/wiki/Continuous_Information_Sensemaking

Topic Tracker

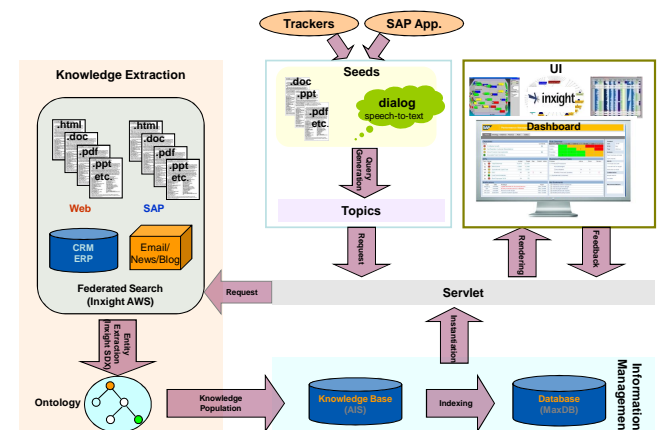
- Topic Tracker is a knowledge discovery and information summarization tool that can **continuously** find, gather, analyze and summarize large amount information using **advanced information extraction, machine learning, and other technologies**, from **statistical and semantic** ways.

► http://research.pal.sap.corp/wiki/Topic_Tracker

Problems

- Process:
 - Automatically extract explicit topics from implicit inputs.
 - Search on the data source.
 - Extract related collections
 - Marked up the topics
 - Summarize amount of information into a readable and intuitive format.
- Heterogeneous data exist everywhere, both inside and outside enterprises.
- Overwhelmed information analysis and understanding
- Valuable information under the water.

Topic Tracker Framework



Topic Tracker v1



Callouts in the screenshot:

- User can upload slides.
- Or directly input seed.
- User can customize either internal or external search sources.
- Content extracted from slides.
- Query candidates generated from content.
- Name entities identified for query generation.
- Query 1
- Federated search results...
- Search result snippet.
- Meta info.
- Search results visualized in Time Wall, Star Tree or Cluster Cloud.

SAP RESEARCH

Statistic Approach



- Data source: appears_at
- Media Types: appears_in
- Classification: appears_as
- Keyword co-occurrence: appears_with

SAP RESEARCH

Scenarios from Situation Tracker



The screenshot shows a list of news items on the left and a network graph of entities on the right. The graph includes nodes for names like Cindy McCain, Buffett, Fey, Oliphant, Greg Brady, Halle Berry, Oliver Stone, Ritchie, Marketing Nurse, KY, Powell, Abrams, John McCain, Elonda Abrams, and Barack Obama.

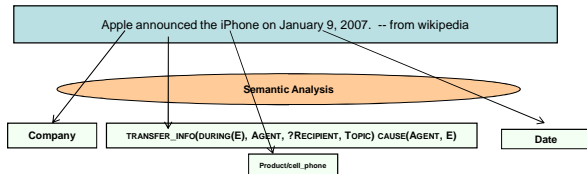
SAP RESEARCH

SAP RESEARCH

The Semantic View of TT



- Identify relations between entities to effectively support decision making and efficiently find information that contextually relevant to users' needs.
- Components:
 - Terms: represent entities, events, actions, qualities, etc. in the domain.
 - Relations: relates entities, qualities, events.



SAP RESEARCH

Definition: Entities & Relationships



- **Entity:** an object or set of objects in the world. Entities may be referenced in a text by their name, indicated by a common noun or noun phrase, or represented by a pronoun.
 - E.g. mentions of a single entity:
 - **Name Mention:** Joe Smith
 - **Nominal Mention:** the guy wearing a blue shirt
 - **Pronoun Mentions:** he, him
 - E.g. Joe Smith, the guy wearing a blue shirt, (PERSON) bought 300 (QUANTITY) shares of SAP AG (ORGANIZATION) in 2006 (DATE)
- Type of Entities: Person, Organization, Location, Time, Facility, ...

SAP RESEARCH

Tools for semantic analysis



- The **relations** are ordered pairs of entities, representing interactions/dependencies between entities.
 - E.g. *George Bush traveled to France on Thursday for a summit.*

Class	Type	Argument1	Argument2
Verbal Asserted Past	PHYS.Located	George Bush	France

SAP RESEARCH

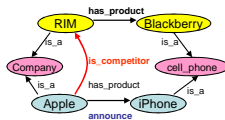
- Database
 - World gazetteer
- Knowledge base
 - Freebase, wikipedia
- NLP tools
 - Inight LinguistX Platform
- Ontology
 - Represent the knowledge, organizing them as a network.

SAP RESEARCH

Application



- Entity Extraction
 - Company/products
- Relation Extraction
- QA
 - What is the product of Apple Inc.?
- Inference



SAP RESEARCH

Conclusion



- **Framework** for continuously find and gather information from large and heterogeneous sources
- Analyze available information from a **statistical** view
- Identify **semantic** relationships between key mentions

SAP RESEARCH

Acknowledge



- Mia Ming Mao
- CIS: Keith, Thomas, Cirrus, Ming, Susanna
- AIS: Horse, Andreas, Max
- Intern family

SAP RESEARCH

Live Demo



Topic Tracker @ SAP

V1.0

SAP RESEARCH



Thank you!