Personalized Semantic Extraction and Retrieval

Arti Ramesh, Zhongfei Zhang, and Weiyi Meng
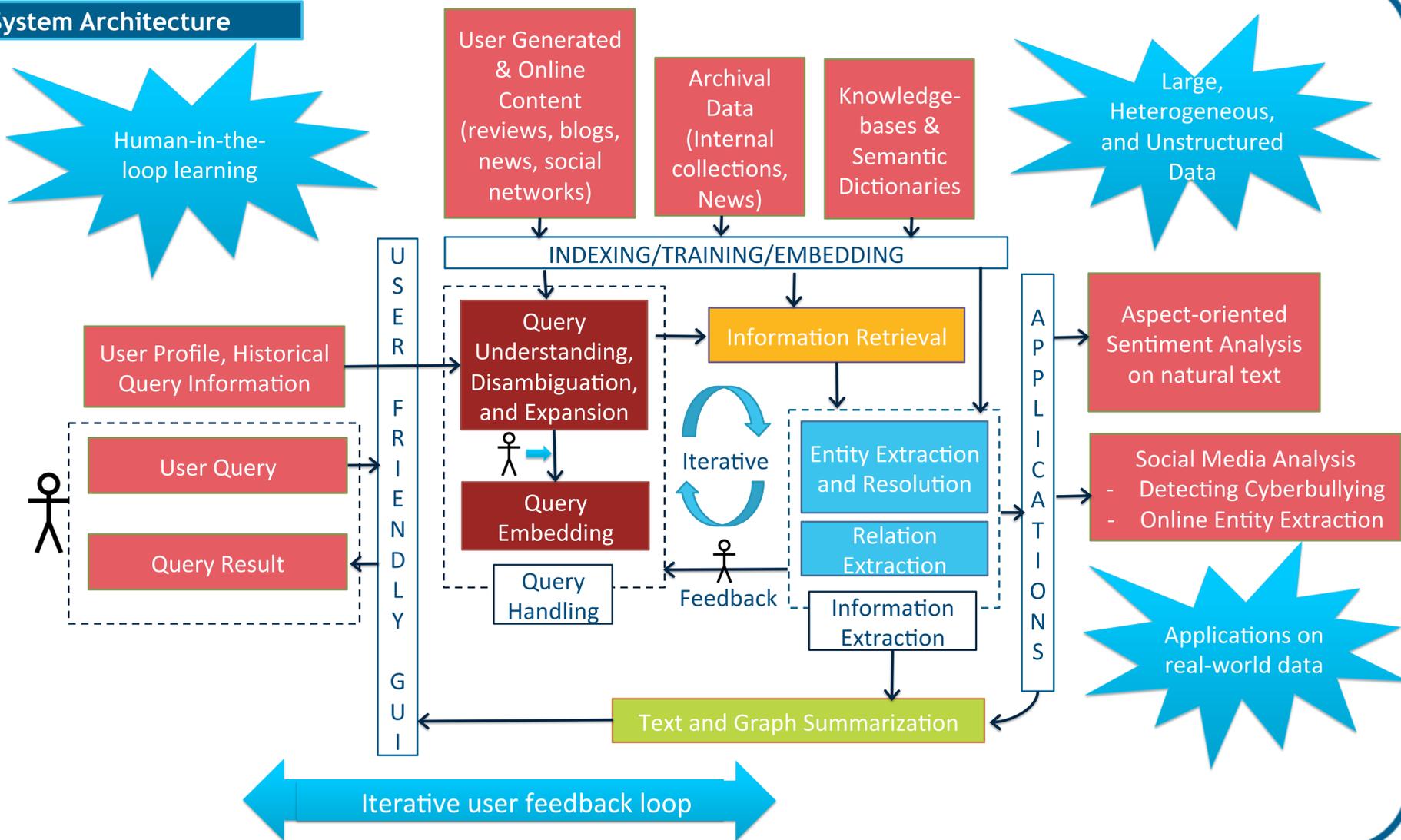
{arti, zzhang, meng} @ binghamton.edu

SUNY Binghamton

Goals and Expertise

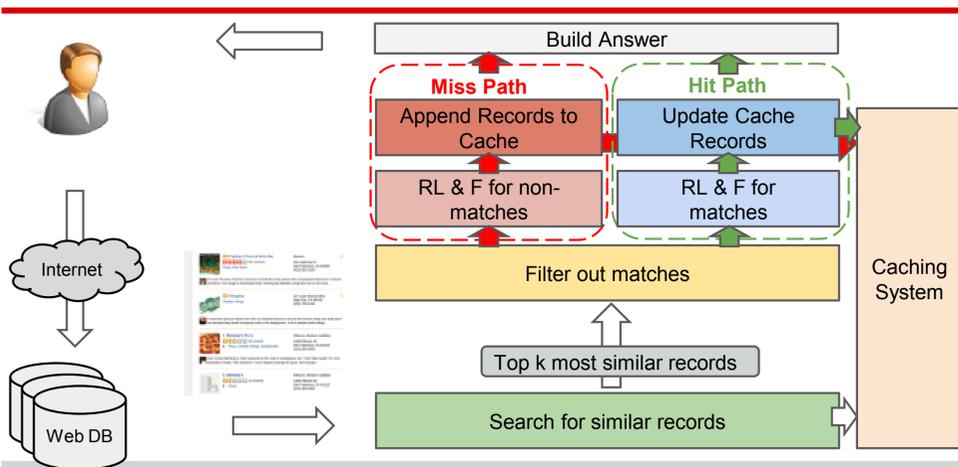Build a user-friendly end-to-end framework for personalized semantic extraction and retrieval
- Information Extraction [Expertise to extract semantics from deep learning, online entity extraction, relation extraction, and entity resolution]
- Information Retrieval
- Human-in-the-loop Interaction
- Applications on diverse real-world problems [Expertise on sentiment analysis, social media and network analysis]

System Architecture

Human-in-the-loop learning

User Generated & Online Content (reviews, blogs, news, social networks)

Archival Data (Internal collections, News)

Knowledge-bases & Semantic Dictionaries

Large, Heterogeneous, and Unstructured Data

USER FRIENDLY GUI

INDEXING/TRAINING/EMBEDDING

User Profile, Historical Query Information

User Query

Query Result

Query Understanding, Disambiguation, and Expansion

Query Embedding

Query Handling

Information Retrieval

Iterative

Entity Extraction and Resolution

Relation Extraction

Information Extraction

Feedback

APPLICATIONS

Aspect-oriented Sentiment Analysis on natural text

Social Media Analysis
- Detecting Cyberbullying
- Online Entity Extraction

Applications on real-world data

Text and Graph Summarization

Iterative user feedback loop

ORLF: A Flexible Framework for Online Record Linkage and Fusion (RL&F)

ORLF Framework

Build Answer

Miss Path
Append Records to Cache
RL & F for non-matches

Hit Path
Update Cache Records
RL & F for matches

Filter out matches

Caching System

Internet

Web DB

Top k most similar records

Search for similar records

Problems

How to efficiently perform RL & F at query time?
- Fast look up using cache indexing

How to avoid repeated RL & F operations?
- Fuse query results with matching records in cache to improve result quality

How to benefit from previous query results to improve RL & F of future queries?
- Use cache to store RL & F results on previous query results

PI Expertise and Key Strengths

Arti Ramesh

Statistical Relational Learning for Heterogeneous Data and real-world applications [AAAI' 14, WI' 17, WWW' 18] Cyberbullying [WWW companion' 18] Aspect-Oriented Sentiment Analysis for natural text [ACL' 15]

Zhongfei (Mark) Zhang

Natural language understanding [AAAI'17] Sentiment Analysis [AAAI'16] Topic Modeling [TKDE'17,TKDE'14,AAAI'10,SIGIR'09,ICDM'09] Relational Data Learning [SDM'15, TKDD'12, AAAI'08, AAAI'07, KDD'07, ICML'07, KDD'06, ICML'06, KDD'05]

Weiyi Meng

Query Understanding and Disambiguation [CIKM'05, CIKM'07, WISE'10, WISE'12], Personalized Search [TKDE'04], IE [NSF BIGDATA, WWW'05, VLDB'06, KDD'07, TKDE'10], IR [SIGIR'04, SIGMOD'07, WISE'12, TOIS'13], Entity Resolution [ICDE'09, ICDE'11, IRI'12]