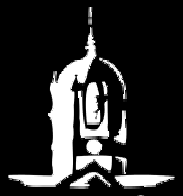
Personalized Question Answering: A Use Case for Business Analysis

VinhTuan Thai[1], Sean O'Riain[2], Brian Davis[1], David O'Sullivan[1]
[1]DERI, National University of Ireland, Galway
[2]Hewlett-Packard, Galway, Ireland

vinhtuan.thai@deri.org

© Copyright 2006 Digital Enterprise Research Institute. All rights reserved.



Outline

- Question Answering Overview
- Personalized Question Answering
- Business Analysis use case
- Personalized Question Answering Framework
- Conclusion and Future work

Question Answering (QA) Overview

- Originated in the 1960s as natural language front ends to databases [1]
- Remains an active research area
- 2 main categories: Open-Domain & Domain-specific QA

Open-Domain QA

- QA track at Text Retrieval Conference TREC [2]
- Pre-defined, large newswire text corpus as knowledge base
- World Wide Web as an auxiliary source of information
- No domain specific knowledge

Domain-specific QA

- Database as knowledge base
- Domain ontologies as knowledge base
- Text collection as knowledge base

Limitations of QA on text collection

- Inadequate consideration of dynamic authoritative source of information
- Inadequate consideration of contextual information
- Ambiguity due to the complex writing style of documents

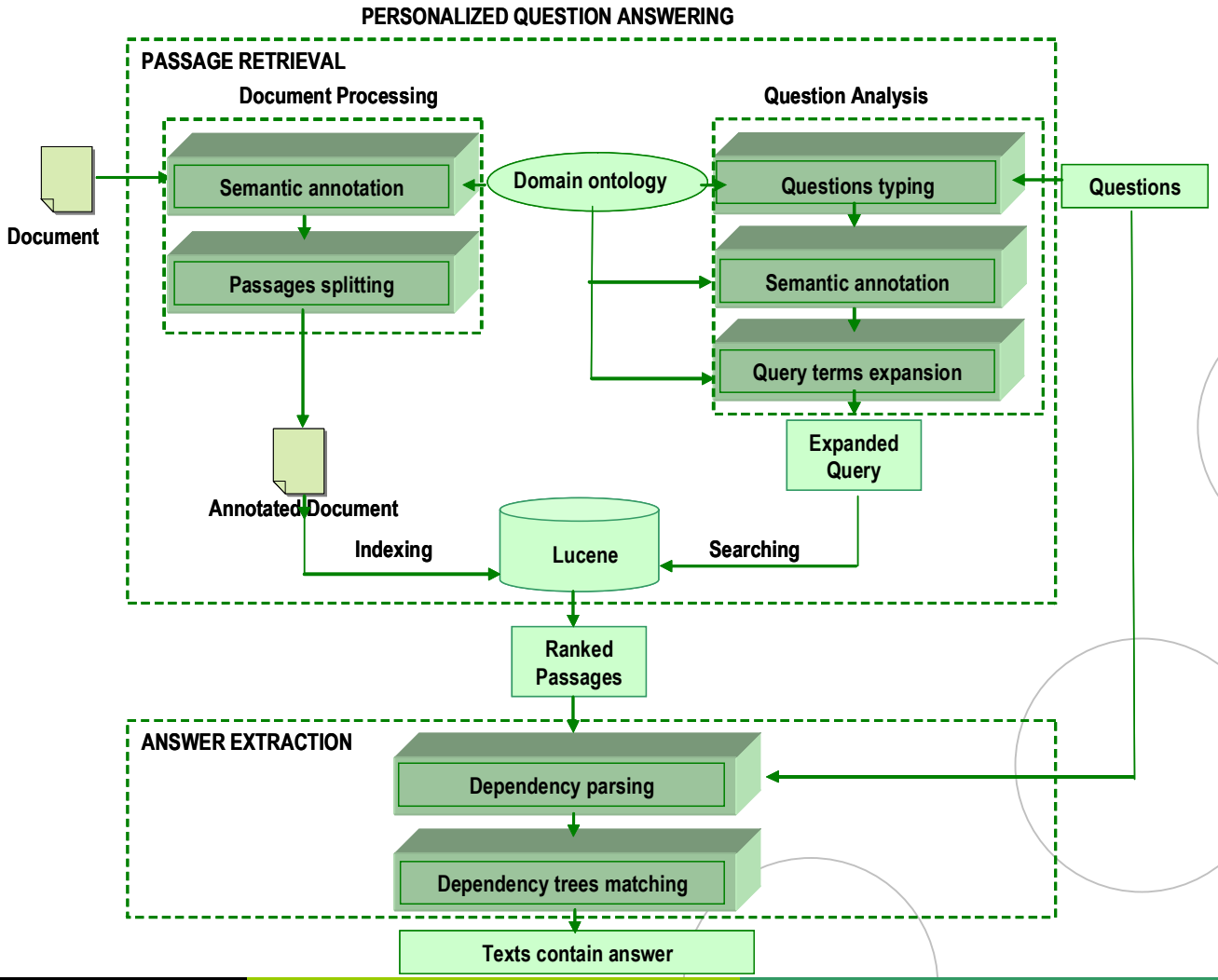
Personalized Question Answering

- domain specific
- dynamic collection of unstructured texts written in rhetorical style as knowledge base
- able to handle various question types
- able to resolve implicit context within questions
- provides an answer-containing chunk of texts rather than the precise answer

Business Analysis use case

- Enterprises perform customer analysis → identify new business opportunities
- Form 10-Q provides financial information and management statements
- Searching, identification & extraction of relevant information → a resource intensive activity
- Potential intuitive and timely solution: QA

Personalized Question Answering Framework



Personalized Question Answering Framework

- Passage Retrieval
  - Document Processing
    - Semantic annotation

      e.g. ”*CompanyX releases a new operating system*.”

      →”*CompanyX* ***BIOntoCompany*** *releases a new operating system* ***BIOntoSoftware***”

      whereby *BIOntoCompany*, *BIOntoSoftware* are labels of ontological concepts *http://localhost/temp/BIOnto#Company*, *http://localhost/temp/BIOnto#Software*

    - Passage splitting & stop-words removal

Personalized Question Answering Framework

- Passage Retrieval
  - Document Indexing
  - Question Analysis
    - Questions typing: pattern-matching rules to map the question type to concepts in the domain ontology

    e.g. (1) *"Which products did CompanyX release?"*

    → *"BIOntoProduct did CompanyX release?"*

    (2) *"Are there any CompanyX's plans to release new products?"*

    → Treated as: *"What are CompanyX's plans to release new products?"*

    - Semantic annotation
    - Query terms expansion: based on sub-class relationships in the domain ontology & synonyms list
  - Searching

- Answer Extraction
  - Why is Answer Extraction necessary?

    e.g. *"Which company acquired Compaq?"*

    → word overlap / term density ranking techniques cannot distinguish *"HP acquired Compaq"* from *"Compaq acquired HP"*

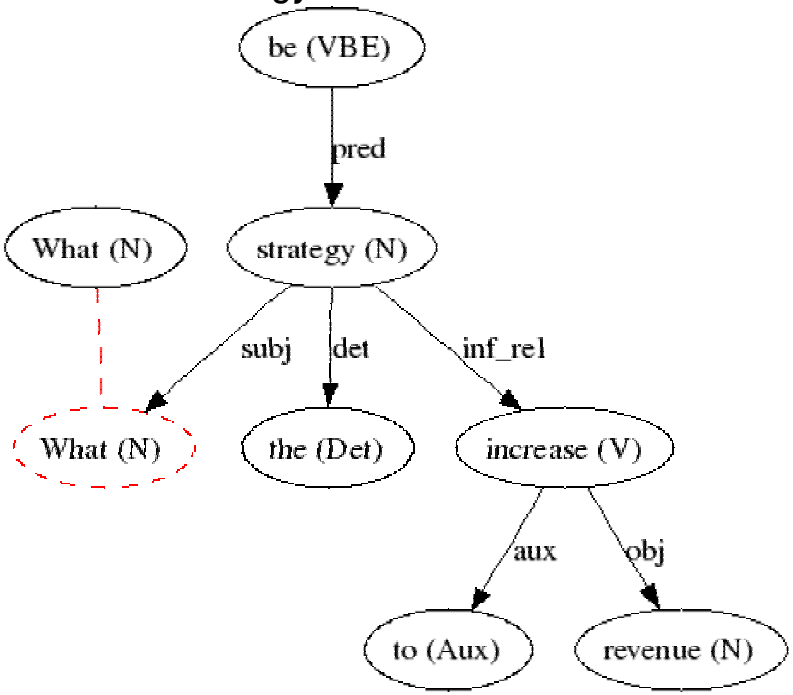
  - Solution: take grammatical constraints / relations of question and candidate sentences into consideration

  - MiniPar Dependency Parser generates dependency trees for words within a given sentence.

Personalized Question Answering Framework

- Answer Extraction
  - Example of dependency tree generated by Minipar visualization tool

"What is the strategy to increase revenue ?"



- Pitfall: strict relations matching suffers substantially from poor recall [3]

- Answer Extraction
  - Potential solution: Approximate/Fuzzy relation matching proposed by Cui et al. [3]
  - Mapping scores model between relation paths based on a variation of a Statistical Translation Model
  - Application to Personalized Question Answering

Conclusion and Future work

- Conclusion
  - Proposed design of Personalized Question Answering framework
  - Business Analysis use case scenario
  - The availability of domain semantics → potential improvement for recall in passage retrieval
  - Approximate dependency matching between question-candidate answer pairs may yield higher precision for answer extraction without impacting on recall
- Future work
  - Examining the possibility of using domain semantics in the Answer Extraction task
  - Handling complex questions whose answers are not explicitly stated
  - Evaluation scheme
  - Integrate the QA system with the Analyst Workbench [4]

References

[1] Hirschman, L., Gaizauskas, R.: Natural language question answering: the view from here. Nat. Lang. Eng. 7 (2001) 275–300

[2] TREC: Text retrieval conference trec http://trec.nist.gov

[3] Cui, H., Sun, R., Li, K., Kan, M.Y., Chua, T.S.: Question answering passage retrieval using dependency relations. In: SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, New York, NY, USA, ACM Press (2005) 400–407

[4] O'Riain, S., Spyns, P.: Enhancing business analysis function with semantics. In Meersma, R., Tari, Z., eds.: On the Move to Meaningful Internet Systems 2006: CoopIS, DOA, GADA and ODBASE; Confederated International Conferences CoopIS, DOA, GADA and ODBASE 2006 Proceedings. LNCS 4275, Springer (2006) 818–835

Thank you