IIIT-Hyderabad @ Guided Summarization Task

System Name: siel

Vasudeva Varma

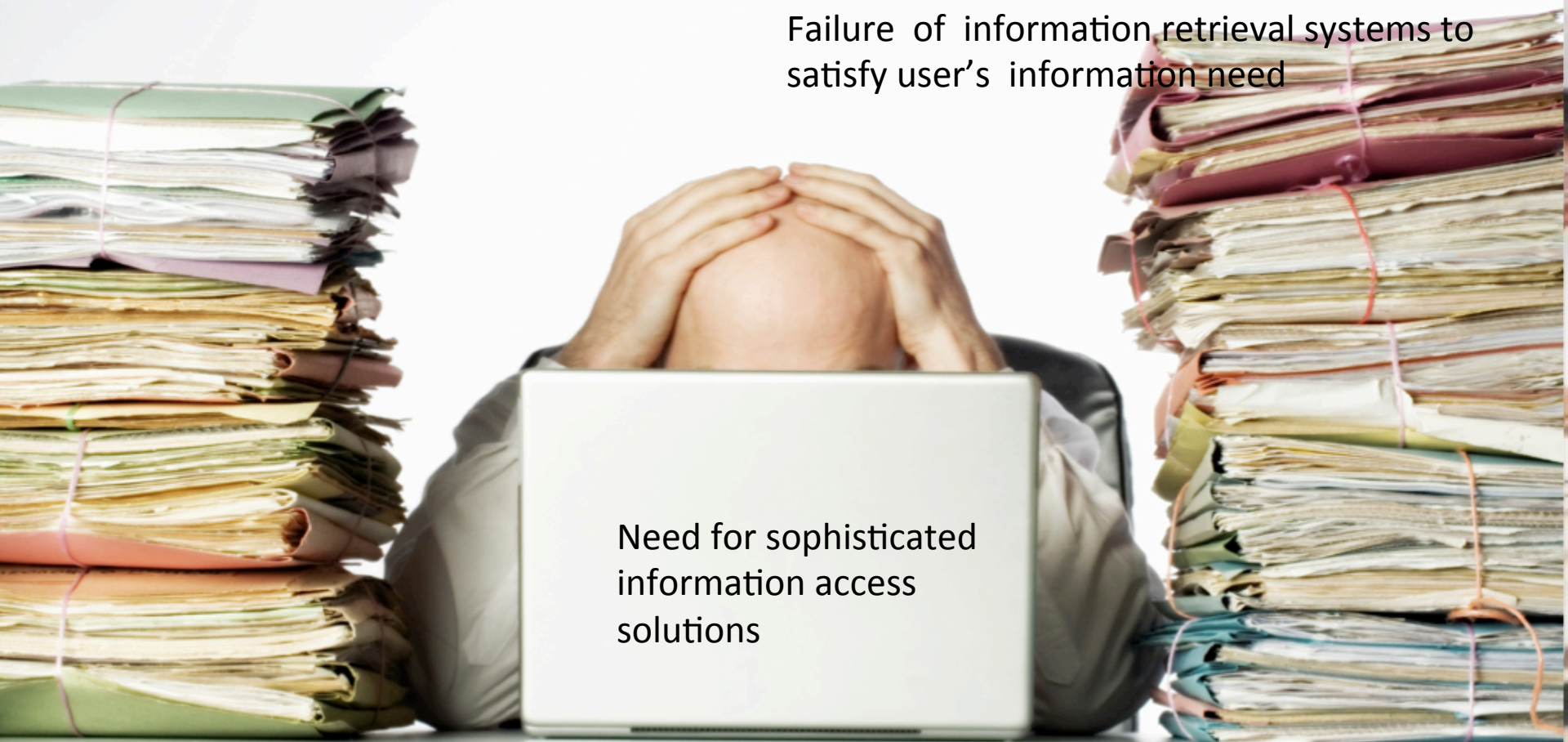**Joint work with Praveen Bysani et al**

Search and Information Extraction Lab

www.iiit.ac.in/~vasu

Information Overload

Explosive growth of information on web

Failure of information retrieval systems to satisfy user's information need

Need for sophisticated information access solutions

Introduction

- ## Query Focused Summarization
  - User's information need expressed as a query along with a narrative
  - Set of documents related to the topic
  - Goal is to produce a shot coherent summary focusing answer to the query

- ## Guided Summarization
  - Each topic is classified into a set of predefined categories
  - Each category has a template of important aspects about the topic
  - Summary is expected to answer all the aspects of template while containing other relevant information

Guided summarization

- Encourage **DEEPER LINGUISTIC AND SEMANTIC KNOWLEDGE** ts
instead of relying only on document word frequencies to select important
concepts

- S **SHARES SIMILARITY WITH INFORMATION EXTRACTION**
  – Specific information from unstructured text is identified and consequently
    classified into a set of semantic labels (templates)
  – Makes information more suitable for other information processing tasks
- A guided summarization system has to produce a readable summary
  encom **INFORMATION ABOUT THE TEMPLATES**
- Very few investigations exploring the potential of merging summarization with

MERGING SUMMARIZATION WITH INFORMATION EXTRACTION

External
Knowledge

Summary

Summarizer

Query

Guided
Summary

When
What
Where
Who
How

IIIT, HYDERABAD
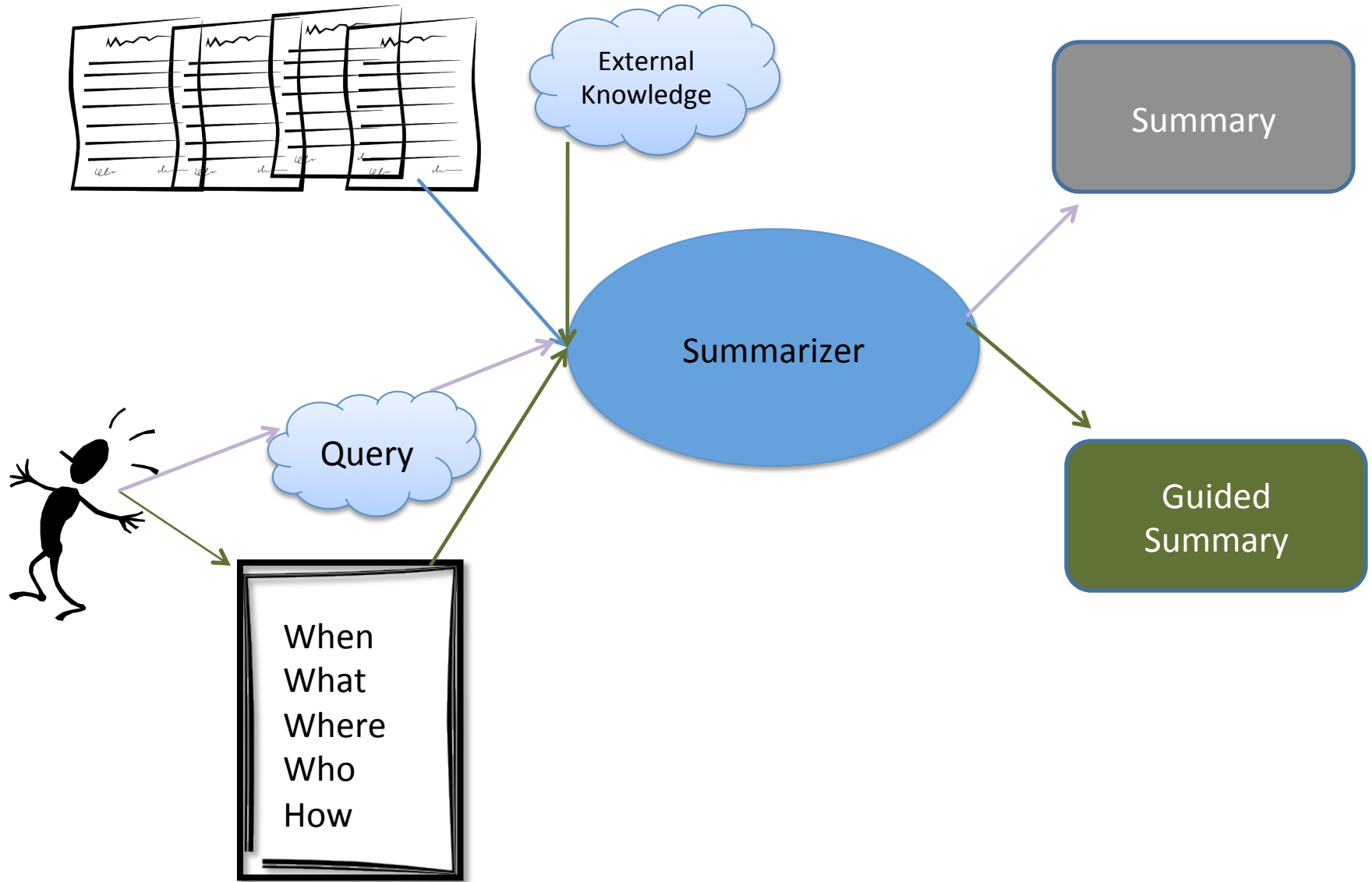
Our Previous System

- ## DUC (2005-2007) and TAC (2008,09)
  - **pHAL**: Probabilistic Hyperspace Analogue to Language
  - **Sentence Prior**

- ## TAC-2009
  - **DFS**: DFS of a word is the ratio of number of docs in which it occurred to total number of docs in cluster
  - **SP**: Sentence Position is a heuristic feature that boosts leading sentences in a document
  - **KL**: Kullback-Leibler Information divergence
  - **Novelty Factor (NF):** Ratio of number of docs in which a word occurred in cluster B to number of docs in which it occurred in cluster A
  - **Regression**: Sentence rank is estimated through support vector regression, from the feature vectors

IIIT, HYDERABAD

Our approach in TAC-2010

- Building a domain model
  - Essential background knowledge for information extraction

- Sentence Annotations
  - To identify sentences having answers to aspects of template

- Concept Mining
  - To use semantic concepts instead of words to calculate sentence importance

- Summary Extraction
  - Modification of summary extraction algorithm to adapt to the requirements using sentence annotations

IIIT, HYDERABAD

Domain model

- Manually collected 500 relevant Wikipedia articles for each category and then extracted first paragraph of each article

THE BHOPAL DISASTER IS THE WORLD'S WORST INDUSTRIAL CATASTROPHE. IT OCCURRED ON THE NIGHT OF DECEMBER 2–3, 1984 AT THE UNION CARBIDE INDIA LIMITED (UCIL)PESTICIDE PLANT IN BHOPAL, MADHYA PRADESH, INDIA. A LEAK OF METHYL ISOCYANATE (MIC) GAS AND OTHER CHEMICALS FROM THE PLANT RESULTED IN THE EXPOSURE OF SEVERAL THOUSANDS OF PEOPLE. ESTIMATES VARY ON THE DEATH TOLL. THE OFFICIAL IMMEDIATE DEATH TOLL WAS 2,259 AND THE GOVERNMENT OF MADHYA PRADESH HAS CONFIRMED A TOTAL OF 3,787 DEATHS RELATED TO THE GAS RELEASE.

- An excerpt from Wikipedia article about Bhopal gas tragedy classified under Accidents and Natural Disasters category
- A multi class classifier built over all the categories using these articles

Sentence annotations

- Sentences in the articles are tags corresponding to aspects in the template

- Template include subjective and objective aspects

- Sentences are annotated manually, as NER would recognize only objective tags

`<WHEN><WHERE>IT OCCURRED ON THE NIGHT OF DECEMBER 2–3, 1984 AT THE UNION CARBIDE INDIA LIMITED (UCIL)PESTICIDE PLANT IN BHOPAL, MADHYA PRADESH, INDIA.</WHEN></WHERE> <HOW>A LEAK OF METHYL ISOCYANATE (MIC) GAS AND OTHER CHEMICALS FROM THE PLANT RESULTED IN THE EXPOSURE OF SEVERAL THOUSANDS OF PEOPLE.</HOW>`

- Annotated sentences are used for building a multi class classifier using simple naïve Bayes classifier

- Summary extraction algorithm is modified to select sentences from ranked list ensuring answers to all the aspects of the template

Concept Mining

- Semantic concepts instead of words to compute sentence importance

- Wikipedia Miner, a freely available toolkit for navigating and making use of content in Wikipedia

- Services to search for entities, comparing two entities and wikifying the snippets of text

- Wikifying the document collection, annotates text with links to the Wikipedia pages

- Semantic relatedness between two concepts using category hierarchy and textual content of the concepts

- Relatedness measure of a concept with all other concepts in document collection as its importance

Role of prepositions

- Preposition is a POS that links nouns, pronouns with other phrases in a sentence
- Represents temporal, spatial, logical relationship of its object to the rest of sentence

Example: The Indian cricket team is travelling **from** England **to** India.

- Prepositions *from, to* convey that there are two entities and a relation between them
- Role of prepositions has never been explored before to estimate sentence importance
- Frequency of a selected set of prepositions in a sentence is measured as its score

$$PrepImp(s) = \frac{\sum_{w_i \in s} IsPrep(w_i)}{|s|}$$

Runs Submitted

- Run 1 (System id:22)
  - DFS + SP + KL for cluster A
  - NF + KL + SP for cluster B

- Run 2 (System id:40)
  - DFS+SP+KL+PrepImp for cluster A
  - NF+KL+SP+PrepImp for cluster B
  - Modified summary extraction algorithm using sentence annotation model

Evaluation and results

Cluster A	ROUGE 2	ROUGE SU4	Pyramid	Responsiveness
Run1	**0.09574 (1/43)**	**0.13014 (1/43)**	**0.425 (1/43)**	3.130 (2/43)
Run 2	0.0695 (23/43)	0.10788 (22/43)	0.347 (21/43)	2.804 (21/43)

Cluster B	ROUGE 2	ROUGE SU4	Pyramid	Responsiveness
Run1	0.06998 (7/43)	0.11107 (4/43)	0.282 (3/43)	2.457 (9/43)
Run 2	0.05894 (23/43)	0.10024 (20/43)	0.207 (26/43)	2.130 (25/43)

At category and aspect level

Category	Pyramid score	Responsiveness
Accidents	0.445	3.429
Attacks	0.524	3.286
Health and safety	0.300	2.583
Endangered Resources	0.396	3.100
Investigations	0.520	3.500

	What	When	Where	Who	Subjective
Accidents	0.621	0.468	0.33	0.457	0.289
Attacks	0.745	--	0.753	0.622	0.491
Health	0.322	--	--	0.172	0.166
Endangered Resources	0.52	--	--		0.128
Investigations	--	--	--	0.785	0.371

Analysis

- Run1 is successful in producing informative summaries for cluster A

- Ranked **first** in all evaluation metrics including pyramids and ROUGE

- Results for update component (cluster B) are satisfying but not on par with cluster A

- Run2 managed to produce summaries that fared near median of all submissions

- Difficulty of task depends on the type of category. Summarizing Health and safety, Endangered resources is relatively hard

- Summaries are able to answer objective aspects with more ease compared to subjective aspects

Post TAC Experiments

Focus of post TAC experiments to improve the quality of update summaries

- HKLID
  - An extension of KL to measure the divergence between Hybrid Language Models (LM) of two sentences that are built over cluster A and cluster B
  - A hybrid language model is the combination of document and sentence language models, for better divergence calculations
  - HKLID between LM's of two sentences s in cluster B and s in cluster A is calculated as,

$$\sum_{w \in s_i} P(w|s_i)P(w|cluster\,B)*log*\frac{P(w|s_i)P(w|cluster\,B)}{P(w|s_j)P(w|cluster\,A)}$$

Discussion

- As language model of Wikipedia is quite different from document collection, it might not have produced desired results

- Answering subjective aspects need more effort compared to objective aspects

- Although the feature PrepImportance did not perform well, we believe that the idea can be used in a more sophisticated way to devise an effective scoring feature

- Post TAC experiments on cluster B resulted in approximate 9% improvement in ROUGE scores over original submission.

	ROUGE-2	ROUGE SU4
Run1	0.0699(7/43)	0.11107(4/43)
System id:16	0.0802(1/43)	0.12006(1/43)
NF+SP+HKLID	0.07899	0.12000
NF+SP+Concept Mining	0.07508	0.11746

Thank You – Questions?

Vasudeva Varma

vv@iiit.ac.in   www.iiit.ac.in/~vasu