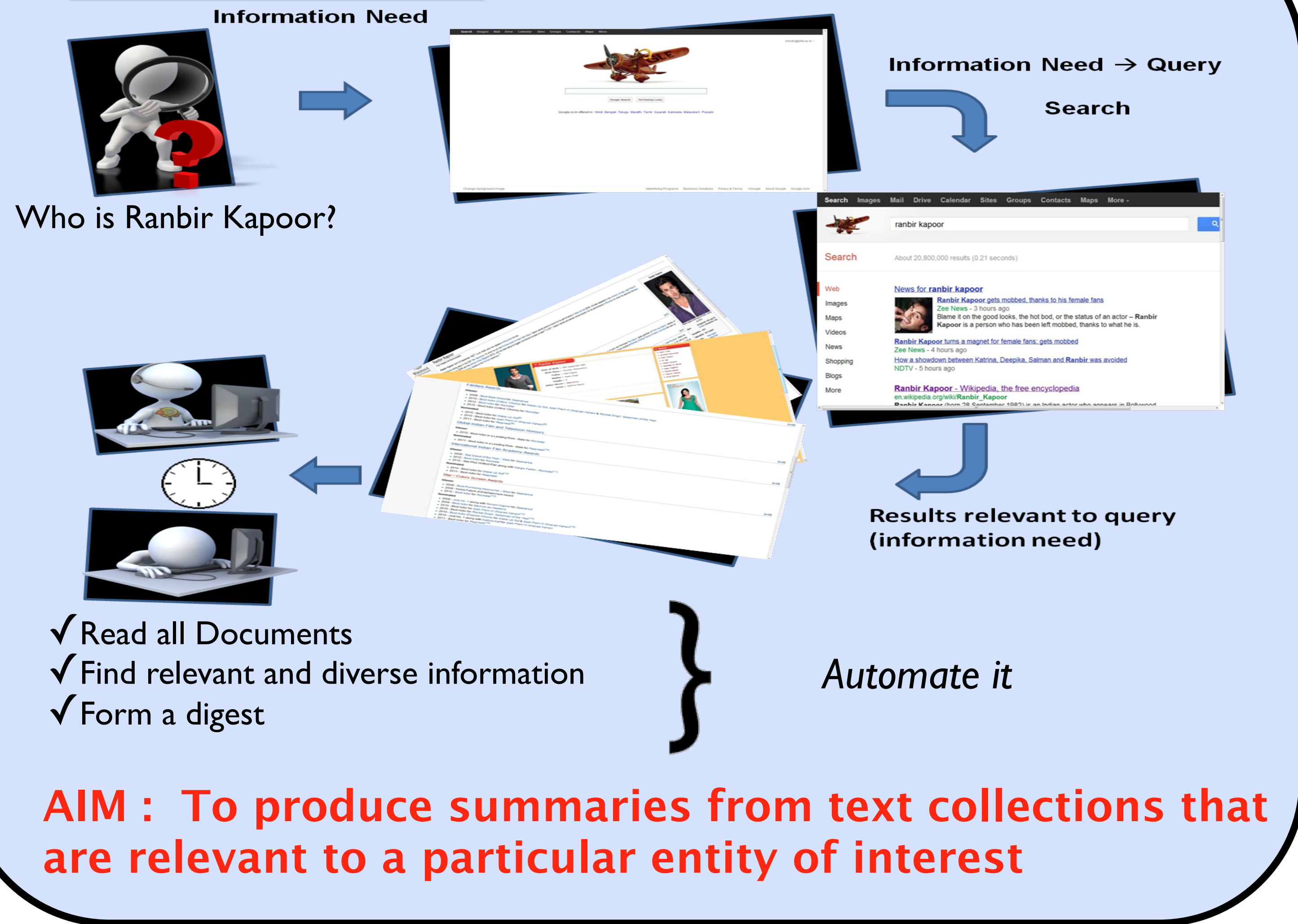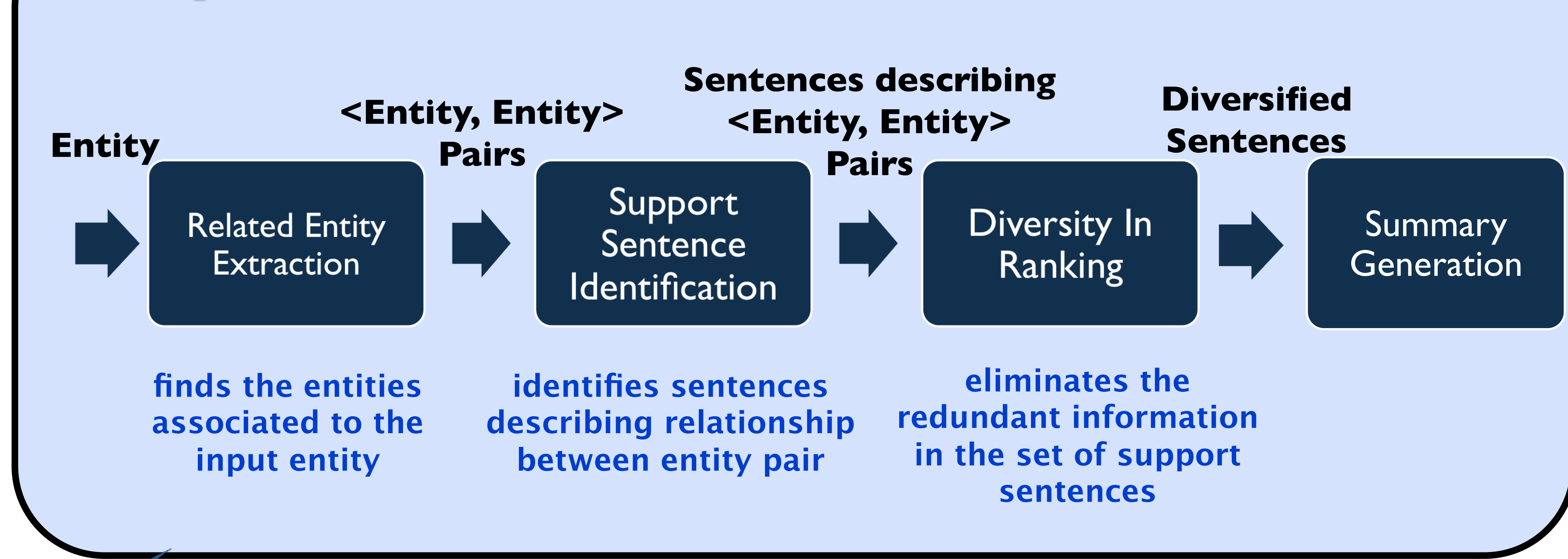# ENTITY-CENTRIC SUMMARIZATION

Shruti Chhabra, Srikanta Bedathur

Indraprastha Institute of Information Technology, Delhi

{shrutic, bedathur}@iiitd.ac.in

Introduction

Information Need

Who is Ranbir Kapoor?

Information Need → Query

Search

Results relevant to query (information need)

✓ Read all Documents
✓ Find relevant and diverse information
✓ Form a digest

} *Automate it*

AIM : To produce summaries from text collections that are relevant to a particular entity of interest

Proposed Framework

Entity → **Related Entity Extraction** → **<Entity, Entity> Pairs** → **Support Sentence Identification** → **Sentences describing <Entity, Entity> Pairs** → **Diversity In Ranking** → **Diversified Sentences** → **Summary Generation**

finds the entities associated to the input entity

identifies sentences describing relationship between entity pair

eliminates the redundant information in the set of support sentences

Illustration

facebook (Organization)

Related Entities
▶ Chris Hughes (Person)
▶ Dustin Moskovitz (Person)
▶ Harvard (Organization)
▶ Facebook (Building)
▶ 2004 (Date)
▶ Mark Zuckerberg (Person)
▶ News Feed (Organzation)
▶ Facebook Platform (Building)
...

Diversified Support Sentences
▶ Zuckerberg launched the online social directory Facebook on February 4, 2004, for Harvard students.
▶ Mark Zuckerberg, Chris Hughes and Dustin Moskovitz (co-founders of Facebook)
▶ Mark Zuckerberg, Facebook creator, b. 1984
▶ Zuckerberg currently works at Facebook's Palo Alto, California headquarters.
▶ The first of the new features , News Feed , appears on every Facebook member 's home page , displaying recent Facebook activities of a member 's friends
...

Related Entities Extraction

Challenges:
◉ Requires the retrieval approach to be domain-independent
◉ Applicable to varied entity types
◉ Retrieved entities must have a relation with the source entity

Support Sentences Identification

Challenges:
◉ Entities mentioned in their various surface forms in document

Proposed Approach:
◉ Entity surface form mapping dictionary
◉ Context plays imp. role
Ranking model :
$$H_{ee'} = F_e(s, C_s) \quad \forall s \in S_{ee'}$$

Diversity In Ranking

Challenges:
◉ Sentences may partially contain overlapping information
◉ Sentences conveying similiar information have different word distribution

References

1. Sauper, C., and Barzilay, R. Automatically generating wikipedia articles: a structure-aware approach. AFNLP (2009), pp. 208–216.

2. Filippova, K., Surdeanu, M., Ciaramita, M., and Zaragoza, H. Company-oriented Extractive Summarization of Financial News. ACL (2009), pp. 246–254.

3. Blanco, R., and Zaragoza, H. Finding Support Sentences for Entities. SIGIR (2010), pp. 339–346