

Building Tempo-IndoWorNet: A Resource for Temporal-IR & NLP

Hindi Wordnet + Time annotation to each synset = Tempo-IndoWordnet



natural language
processing

Objective

Annotate temporal information to each synset of Hindi wordnet

'e.g. i.

▮ Word--> पत्र -

▮ Synset-> पत्ता, पात, पर्ण, पत्र, दल, छद, पत्रक, परन, वर्ह; पेड़-

▮ Gloss--> पौधों में होने वाला विशेषकर हरे रंग का वह पतला, हल्का अवयव जो उसकी टहनियों से निकलता है

▮ Temporal annotation--> Atemporal



natural language
processing

b.

Word --> कल -

Synset--> भविष्य, भविष्य काल, आगामी समय, उत्तरकाल, उत्तर-काल, उत्तर काल, भावी समय, आने वाला समय, अगत, अप्राप्तकाल, अवर्तमान, अवर्तमान, आगम, आगाह;

Gloss--> आने वाला काल या समय;

Example sentence--> "भविष्य में क्या होगा कोई नहीं जानता । / कल किसने देखा है ।"

Temporal annotation--> Future

c.

Word--> अब;

Gloss--> इस समय (विषय या कार्य परिवर्तन सूचक) ;

Example sentence--> "अब अगली समस्या पर विचार करें ।"

Temporal Annotation--> Present



natural language
processing

Challenges

▮ Polysemous words with different temporal annotations

▮ e.g.

▮ a. word- कल

▮ Synset- अतीत, अतीतकाल, भूतकाल, अतीत काल, कल, भूत काल, गत काल,
▮ पिछला ज़माना, पूर्वकाल;

▮ Gloss- बीता हुआ समय या काल;

▮ Example Sentence- "यह उपन्यास अतीत की घटनाओं पर आधारित है । / कल
▮ की बातों को याद करके दुखी होना अच्छा नहीं ।"

▮ Temporal Annotation- Past

▮ b. word- कल

▮ Synset- कल;

▮ Gloss- आज के बाद आनेवाले पहले दिन को ;

▮ Example sentence- "मैं कल घर जाऊँगा ।"

▮ Temporal Annotation- Future



natural language
processing

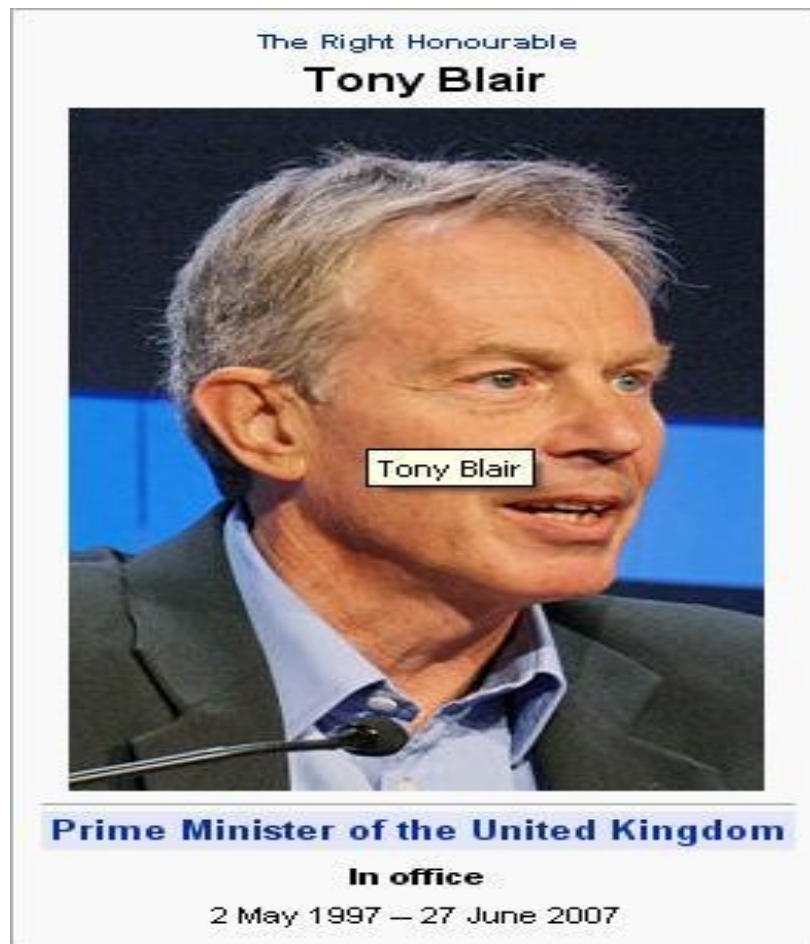
Motivation

Temporal Information Retrieval



natural language
processing

Example of Temporal-IR



query: PM United Kingdom 2000

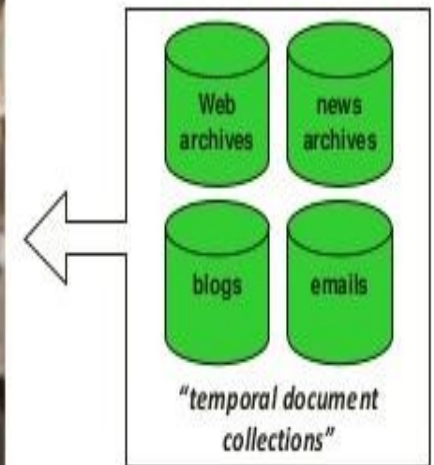
With T-IR United kingdom's PM during 2000 will be returned.



natural language
processing

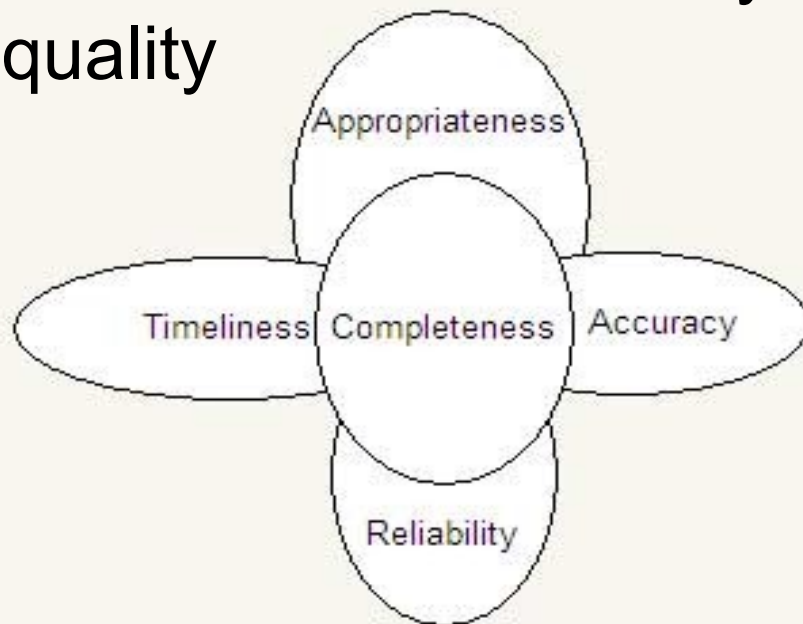
Why temporal information retrieval?

>Value of info is time dependent



Term-based IR approaches
may give unsatisfied results

>Time is one of the key 5 aspect to determine document's quality



natural language
processing

NLP & IR lacks temporal lexical resources.

Document Creation Time: 1998-04-18
Hungarian astronaut Bertalan Farkas is leaving for the
United States to start a new career, he said today.
... On May 22, 1995, Farkas was made a brigadier general,
and the following year he was appointed military attache
... However, cited by District of Columbia traffic police in
December for driving under the influence of ...

Figure 1: Examples of temporal expressions in a news article with explicit (transparent boxes) and relative (solid boxes) expressions. Arrows indicate what kind of context information is needed to normalize the temporal expression.



natural language
processing

Research trends in Temporal Information Retrieval

- i. Exploratory search
- ii. Micro-blogging & real time search
- iii. Temporal clustering
- iv. Temporal Querying
- v. Temporal Summaries , etc



Goal

Temporal Resource such as Tempo-IndoWordNet would

- i. facilitate research in T-IR & may even give rise to new research trends In T-IR & NLP just as wordNet facilitated in NLP
- ii. Bridge the gap when no explicit time mentions are present



natural language
processing

Method

- i. Bootstrapping
 - a. Start with a seed entities
 - b. Classify HindiWordNet synsets
- ii. Classification strategies
 - One-step
 - Two-step
- iii. Classifier used
 - Naive-Bayes



natural language
processing

Work done so far

- i. So far we have worked on one step classification & as expected atemporality is spreading a lot in it
- ii. Now we are working on two step classification.

THANK YOU
ANY QUESTIONS

