



Visvesvaraya Technological University Belgaum

## **KANNADA WORD SENSE DISAMBIGUATION (Kannada WSD)**

---

**Dr. V.N. Narayana**

**Professor and Head**

**S. Parameswarappa**

**Research scholar**

**Department Of Computer Science  
Malnad College Of Engineering  
Hassan-573 201  
[www.mcehassan.ac.in](http://www.mcehassan.ac.in)**

**[vnn@mcehassan.ac.in](mailto:vnn@mcehassan.ac.in)  
[sp\\_gpt@yahoo.co.in](mailto:sp_gpt@yahoo.co.in)**

# Overview

---

## Kannada WSD

- **Introduction**
- **Motivation**
- **Problem Definition**
- **Approaches to WSD**
- **Kannada Wordnet**
- **Methodology**
- **Evaluation**
- **Conclusion**
- **References**
- **Discussion**

# Introduction

## Kannada WSD

- Word Sense Disambiguation (WSD) is defined as the task of finding the correct sense of a word in a specific context.
- Words can have different senses. Some words may have multiple meanings. This is called Polysemy.  
Example: ಅದಿರು [adhiru] can be a 'Raw-Material' or 'Shaking'.
- Word sense disambiguation is the problem of determining in which sense a word having a number of distinct senses is used in a given sentence.

## Introduction (Cont...)

### Kannada WSD

Consider the following sentence

1. ರಾಮನ ಮನೆಯ ನೀರೆತ್ತುವ ಯಂತ್ರ ಹಾಳಾಗಿದೆ.  
[Ramana maneya neereththuva yanthra halagide]  
'Ramas house water lifting machine is got corrupted'
2. ರಾಮನು ಪಂಡಿತರ ಬಳಿ ಹೋಗಿ ಯಂತ್ರ ಹಾಕಿಸಿಕೊಂಡು ಬಂದನು.  
[Ramanu pandithara bali hogi yanthra hakisikondu bandanu]  
'Ramu received an astrological device from the astrologer'.

# Motivation

---

## Kannada WSD

- Word sense disambiguation, a task of removing the ambiguity of a word in a context, is important for many Natural Language Processing applications such as
  1. Information Retrieval.
  2. Machine Translation.
  3. Part of Speech tagging.

# Problem Definition

## Kannada WSD

- We propose an approach to solve a Kannada word sense ambiguity problem using the Kannada Wordnet.
- The Word Sense Disambiguation task involves three steps.
  1. The first step is to determine all the different senses for every word relevant to the text under consideration.
  2. The second step involves a means to assign the appropriate sense to each occurrence of a Word in a context.
  3. In the third step, the computer needs to learn how to associate a word sense with a word in a context.

# Approaches to WSD

---

## Kannada WSD

There are three conventional approaches to WSD.

1. Dictionary and knowledge based methods.
2. Supervised methods.
3. Unsupervised methods.

# Kannada Wordnet

---

## Kannada WSD

- ❑ Kannada Wordnet is an on-line lexical reference system for Kannada language whose design is inspired by the English and Hindi Wordnet.
- ❑ Kannada nouns, verbs, adjectives and adverbs are organized into synonym sets (synsets), each representing one underlying concept.
- ❑ Different relations link the synonym sets,
- ❑ Wordnet being organized according to semantic relations.



# Kannada Wordnet – System Design

---

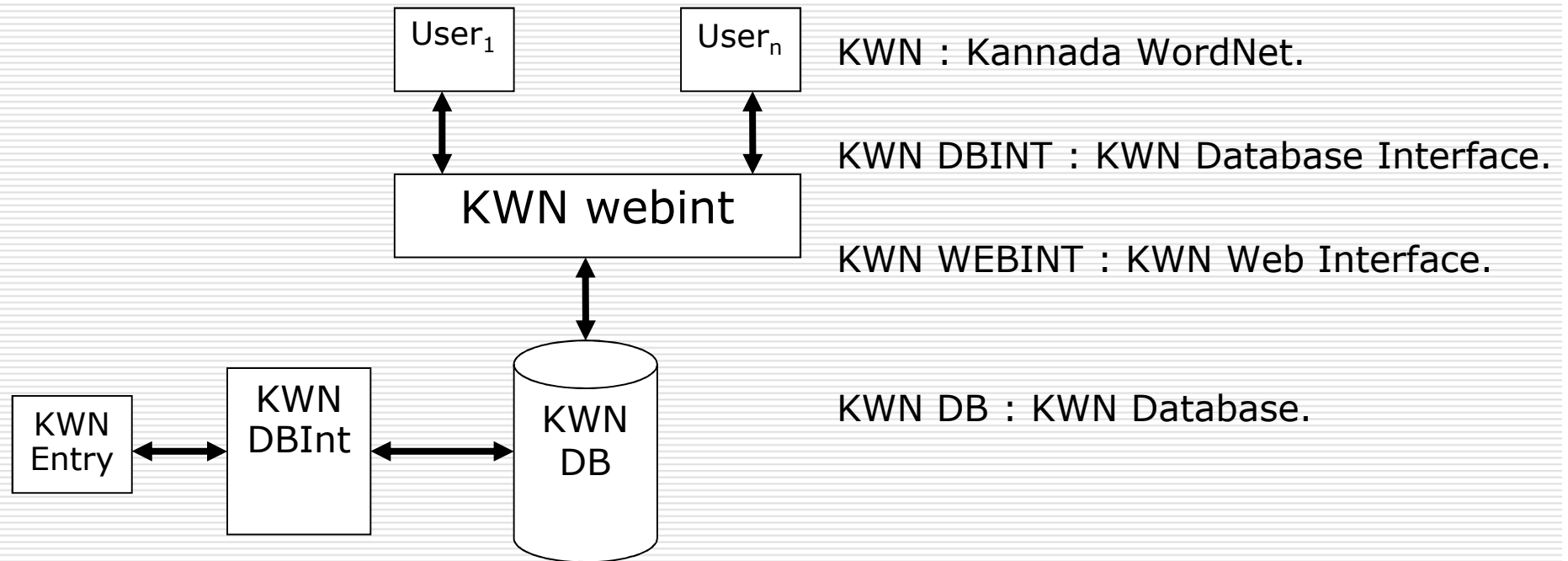
## Kannada WSD

- The Kannada Wordnet system basically consists of two parts
  - 1.lexical interface: The lexical interface is where the lexical entry of a word is entered by the lexicographer.
  - 2.web interface: The web interface provides provision for the user to search the word through the query, and the result is displayed in readable form.

## Kannada Wordnet – System Design(Cont...)

### Kannada WSD

The system architecture can be shown as follows



# Kannada Wordnet – Database

---

## Kannada WSD

- ❑ The lexical database is stored in a RDBMS.
- ❑ The synsets and semantic links between them are grouped according to different grammatical categories like noun, adjective, verb and adverb and are stored in a separate tables as indicated in the Table 1.
- ❑ Each column heading indicates the name of the table and column entries indicate the fields.
- ❑ An instance of the Kannada WordNet database, where a synset is entered is shown in Table 2.
- ❑ Synset-ID is the unique number that is associated with the Synset.

## Kannada Wordnet – Database(Cont...)

### Kannada WSD

Table 1. Tables of Synsets and semantic relations in the database.

Synset table	Noun table	Verb table	Adverb table	Adjective table
Synset-Id Head-word Synset Gloss POS	Synset-Id Hypernyms Antonyms Hyponyms Meronyms Holonys	Synset-Id Hypernyms Antonyms Hyponyms Entailment Troponymy Complex Verbs	Synset-Id Antonyms	Synset-Id Antonyms

## Kannada Wordnet – Database(Cont.)

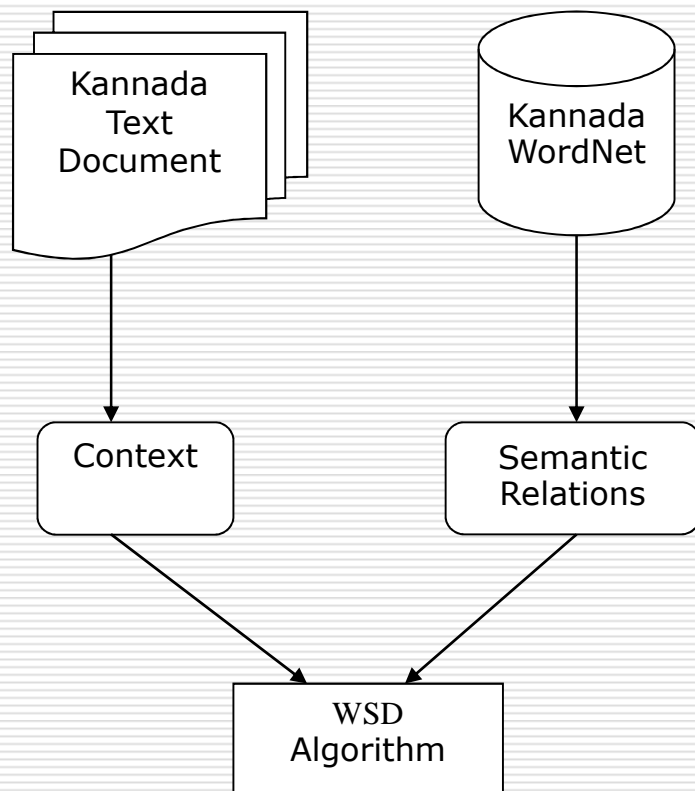
### Kannada WSD

Table 2. Sample records from the Kannada Wordnet database.

Synset-Id	Synset	Gloss	POS
1112	ಲೇಖನ, ಬರವಣಿಗೆ	ಬರಹಗಾರರು ಬರೆಯುವ ಗ್ರಂಥ: ರಾಮಯಣ ದರ್ಶನಂ ಲೇಖನವನ್ನು ಬರೆದವರು ಕುವೆಂಪು.	Noun
1121	ಕೂಡಿಸು, ಸೇರಿಸು , ಜೊತೆಮಾಡು, ಸಂಗ್ರಹಿಸು	ಸಂಗ್ರಹಿಸಿಡುವ ಕ್ರಿಯೆ : ರಾಮನಿಗೆ ಅಂಜಿ ಚೀಟಿ ಸಂಗ್ರಹಿಸುವುದು ಒಂದು ಹವ್ಯಾಸ.	Verb

# Methodology

## Kannada WSD



- ❑ We describe a Dictionary and Knowledge based approach for assigning senses to words in Kannada.
- ❑ A word is assigned a sense with the use of
  - (i) the context in which it has been mentioned.
  - (ii) the information in the Kannada Wordnet
  - (iii) the overlap between these two pieces of information.
- ❑ The sense with the maximum overlap is the winner sense.

## Evaluation - Test Document

### Kannada WSD

Determine the sense of the word ನೆರೆ [Nere] in the following Kannada text.

ರಾಯಚೂರಿನ ಭೀಮಾನಾಯ್ಕರ ಮಗಳು ರಮಾಬಾಯಿ ನೆರೆದ ದಿನ, ಅಪಾರ ಮಳೆಯಿಂದ ಕೃಷ್ಣ, ತುಂಗಭದ್ರ ನದಿಗಳೆಲ್ಲ ತುಂಬಿ ಹರಿಯಲಾರಂಭಿಸಿ ಪ್ರವಾಹ ಉಂಟಾಗಿ ನೆರೆ ಬಂದಿದ್ದರಿಂದ, ಉತ್ತರ ಕರ್ನಾಟಕದಾದ್ಯಂತ ಜನ ಜಾನುವಾರುಗಳೆಲ್ಲ ಸಂಕಷ್ಟಕ್ಕೀಡಾದವು. ನೆರೆ ಉಂಟಾಗಿದ್ದರಿಂದ, ನೆರೆಹೊರೆಯವರನ್ನೆಲ್ಲಾ ಕರೆದು ರಮಾಬಾಯಿಯವರ ನೆರೆದ ಕಾರ್ಯಕ್ರಮವನ್ನು ಮಾಡಬೇಕೆಂದು ತೀರ್ಮಾನಿಸಿದ್ದ ಭೀಮಾನಾಯ್ಕರು ಸದರಿ ಕಾರ್ಯಕ್ರಮವನ್ನು ರದ್ದು ಮಾಡಿದರು. ನೆರೆ ಬಂದ ದಿನ, ಸುಮಾರು ೮೦ ವರ್ಷದ ನೆರೆ ಬಂದ ಮುದುಕ ರಮಾನಾಯ್ಕ ಎಂಬುವರು ತುಂಗಭದ್ರ ನದಿಯ ನೆರೆ ಯಲ್ಲಿ ಸಿಕ್ಕಿಕೊಂಡಿದ್ದರು. ರಾಯಚೂರಿನ ಜನರೆಲ್ಲ ನೆರೆದು ರಮಾನಾಯ್ಕರನ್ನು ಬದುಕಿಸಿದರು.

## Evaluation - Test Document (Cont...)

### Kannada WSD

#### Kannada Text Transliteration

[Raichurina Bheemanaikara magalu Ramabai Nereda dina, aapara male inda Krishna, Thungabadra nadigalella thumbi hariyalaarambisi pravaha untagi Nere bandiddarinda Uttara karnatakadaadyantha jana jaanuvarugalella sankastakkidadavu. Nere untagiddarinda, Nerehoreyavarennella karedu Ramabaiyavara Nereda karyakramavannu madabekendu theermanisidda Bheemanaikaru sadari karyakramavannu raddu madidaru. Nere banda dina, sumarau 80 varshada Nere banda muduka Ramanaika embuvaru Thungabadra nadiya Nereyalli sikkikondaru. Raichurina janarella Neredu Ramanaikanannu badukisidaru].



## Evaluation - Test Document (Cont...)

### Kannada WSD

#### English\_Translation

The day on which a Raichur resident Bheemanaik's daughter Ramabai got matured (biologically), due to heavy rain, the rivers Krishna and Thungabadra flew heavily and end up in causing floods, the lives of entire Uttara Karnataka's people and animals set into trouble.

Due to flood, Bheemanaik, who has decided to celebrate Ramabai's matured event by calling neighbors has stopped the event. On the same flood day, 80 years white haired old man Ramanaik struck in Thungabadra river flood. All Raichur's people gathered and saved Ramanaik's life.

**WELCOME TO Kannada Word Net - Microsoft Internet Explorer**

File Edit View Favorites Tools Help

Address http://localhost:8080/KannadaWordNet/Display.action Go

**Determine the sense of:**

the word "ನೆರೆ"

**in the following text:**

ರಾಯಚೂರಿನ ಭೀಮಾನಾಯ್ಕರ ಮಗಳು ರಮಾಬಾಯಿ ನೆರೆದ ದಿನ, ಅಪಾರ ಮಳೆಯಿಂದ ಕೃಷ್ಣ, ತುಂಗಭದ್ರ ನದಿಗಳೆಲ್ಲ ತುಂಬಿ ಹರಿಯಲರಂಭಿಸಿ ಪ್ರವಾಹ ಉಂಟಾಗಿ ನೆರೆ ಬಂದಿದ್ದರಿಂದ ಉತ್ತರ ಕರ್ನಾಟಕದಾದ್ಯಂತ ಜನ ಜಾನುವಾರಗಳೆಲ್ಲ ಸಂಕಷ್ಟಕ್ಕೀಡಾದವು. ನೆರೆ ಉಂಟಾಗಿದ್ದರಿಂದ, ನೆರೆಹೊರೆಯವರನ್ನೆಲ್ಲಾ ಕರೆದು ರಮಾಬಾಯಿಯವರ ನೆರೆದ ಕಾರ್ಯಕ್ರಮವನ್ನು ಮಾಡಬೇಕೆಂದು ತೀರ್ಮಾನಿಸಿದ್ದ, ಭೀಮಾನಾಯ್ಕರು ಸದರಿ ಕಾರ್ಯಕ್ರಮವನ್ನು ರದ್ದು ಮಾಡಿದರು. ನೆರೆ ಬಂದ ದಿನ, ಸುಮಾರು ೮ಂ ವರ್ಷದ ನೆರೆ ಬಂದ ಮುರುಕ ರಮಾನಾಯ್ಕ ಎಂಬವರು ತುಂಗಭದ್ರ ನದಿಯ ನೆರೆ ಯಲ್ಲಿ ಸಿಕ್ಕಿಕೊಂಡಿದ್ದರು. ರಾಯಚೂರಿನ ಜನರೆಲ್ಲ ನೆರೆದು ರಮಾನಾಯ್ಕರನ್ನು ಬದುಕಿಸಿದರು.

Analyze

ರಾಯಚೂರಿನ ಭೀಮಾನಾಯ್ಕರ ಮಗಳು ರಮಾಬಾಯಿ ನೆರೆ/1[ಪ್ರಾಯಕ್ಕೆ ಬರು] ದ ದಿನ, ಅಪಾರ ಮಳೆಯಿಂದ ಕೃಷ್ಣ, ತುಂಗಭದ್ರ ನದಿಗಳೆಲ್ಲ ತುಂಬಿ ಹರಿಯಲರಂಭಿಸಿ ಪ್ರವಾಹ ಉಂಟಾಗಿ ನೆರೆ/2[ಪ್ರವಾಹ ಅತಿರುಷ್ಟಿ] ಬಂದಿದ್ದರಿಂದ ಉತ್ತರ ಕರ್ನಾಟಕದಾದ್ಯಂತ ಜನ ಜಾನುವಾರಗಳೆಲ್ಲ ಸಂಕಷ್ಟಕ್ಕೀಡಾದವು. ನೆರೆ/2[ಪ್ರವಾಹ ಅತಿರುಷ್ಟಿ] ಉಂಟಾಗಿದ್ದರಿಂದ, ನೆರೆ/3[ಅಕ್ಕವಕ್ಕ] ಹೊರೆಯವರನ್ನೆಲ್ಲಾ ಕರೆದು ರಮಾಬಾಯಿಯವರ ನೆರೆ/1[ಪ್ರಾಯಕ್ಕೆ ಬರು] ದ ಕಾರ್ಯಕ್ರಮವನ್ನು ಮಾಡಬೇಕೆಂದು ತೀರ್ಮಾನಿಸಿದ್ದ, ಭೀಮಾನಾಯ್ಕರು ಸದರಿ ಕಾರ್ಯಕ್ರಮವನ್ನು ರದ್ದು ಮಾಡಿದರು. ನೆರೆ/2[ಪ್ರವಾಹ ಅತಿರುಷ್ಟಿ] ಬಂದ ದಿನ, ಸುಮಾರು ೮ಂ ವರ್ಷದ ನೆರೆ/4[ಕೂದಲು ಬೆಳ್ಳೆಗಾಗು] ಬಂದ ಮುರುಕ ರಮಾನಾಯ್ಕ ಎಂಬವರು ತುಂಗಭದ್ರ ನದಿಯ ನೆರೆ/2[ಪ್ರವಾಹ ಅತಿರುಷ್ಟಿ] ಯಲ್ಲಿ ಸಿಕ್ಕಿಕೊಂಡಿದ್ದರು. ರಾಯಚೂರಿನ ಜನರೆಲ್ಲ ನೆರೆ/5[ಗುಂಪುನೇರು] ದು ರಮಾನಾಯ್ಕರನ್ನು ಬದುಕಿಸಿದರು.

**Definations:**

Sense	Defination
1	ಪ್ರಾಯಕ್ಕೆ ಬರು
2	ಪ್ರವಾಹ ಅತಿರುಷ್ಟಿ
3	ಅಕ್ಕವಕ್ಕ
4	ಕೂದಲು ಬೆಳ್ಳೆಗಾಗು
5	ಗುಂಪುನೇರು

Copyright © 2010 WSD.com

Done Local intranet

# Conclusion

## Kannada WSD

- We have surveyed various supervised, unsupervised, dictionary based approaches as a solution for Word Sense Disambiguation. After survey or initial overview we would like to do the following:
  1. Analyze and evaluate existing supervised, unsupervised and dictionary based approaches to know about the strengths and shortcomings of the existing systems.
  2. Improvement in earlier proposed solutions, if possible.
  3. Investigate dictionary based approaches in detail.
  4. Explore other approaches.
  5. Building Kannada Wordnet.
- The final aim of this research is to develop a solution for Kannada word sense disambiguation problem.

## References

---

### Kannada WSD

- ❑ Esha Palta, Prof. Om Damani. 2007. Word Sense Disambiguation: First Stage Report. Kanwal Rekhi School of Information Technology. Indian Institute of Technology, Powai, Mumbai.
- ❑ Fellbaum Christiane. 1998. WordNet: An electronic Lexical database. MIT Press.
- ❑ Kalyanamalini Sahoo , Eshwarchandra Vidyasagar V. 2003. Kannada WordNet - a lexical database. TENCON. Conference on Convergent Technologies for Asia-Pacific Region. 4.1352 – 1356.
- ❑ Prabhakar Pande, Lakshmi Kashyap, Manish Sinha, Mahesh Kumar and Pushpak Bhattacharyya. November 2004. Hindi word sense disambiguation. In International Symposium on Machine Translation, Natural Language Processing and Translation Support Systems, Delhi, India.

# Discussion

---

**Kannada WSD**

Q & A

# KANNADA WORD SENSE DISAMBIGUATION

Kannada WSD

Thank you



Dr. VNN & SP