

BWN - A Software Platform for Developing Bengali WordNet



Farhana Faruqe
Mumit Khan

Center for Research on
CRBLP
Bangla Language Processing

Center for Research on Bangla Language Processing (CRBLP),
BRAC University, Dhaka, Bangladesh.

Road Map

- What is WordNet ?
- What is BWN?
- Design and Implementation
 - Lexical Source File
 - Grinder
 - WordNet Database
 - Interface
- Conclusion and Future Plan

What is WordNet?

- A lexical semantic database
- The basic building block of WordNet
- Synset relation in Wordnet
- 61 WordNets in 49 languages in the world

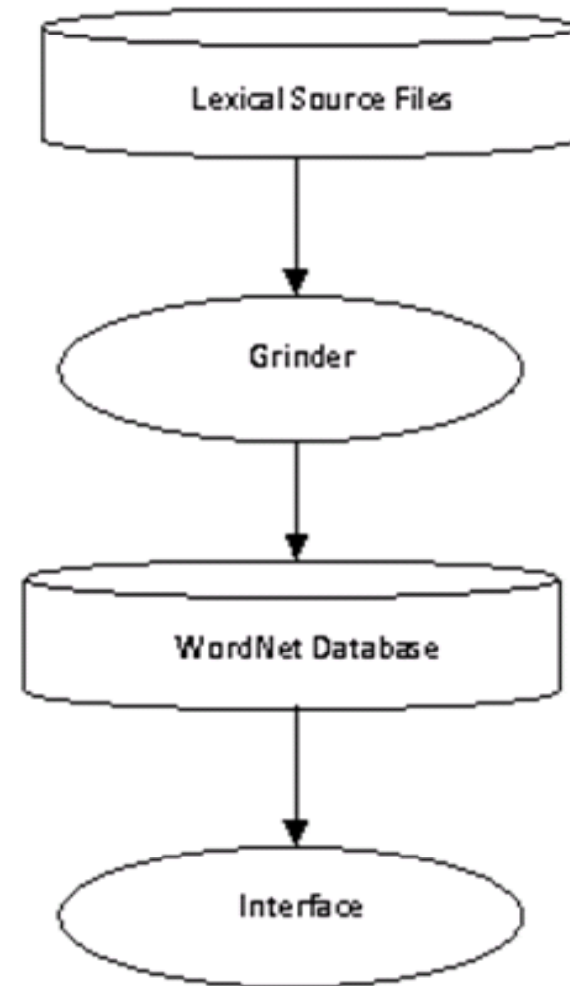
http://www.globalwordnet.org/gwa/wordnet_table.htm

What is BWN ?

- BWN, a software framework to build and maintain a Bengali WordNet
- Two primary challenges –
 - *Create the lexical data*
 - *Build the software framework*
- Two key features of BWN
 - *Incrementally building the WordNet*
 - *Editing the data using a visual interface*

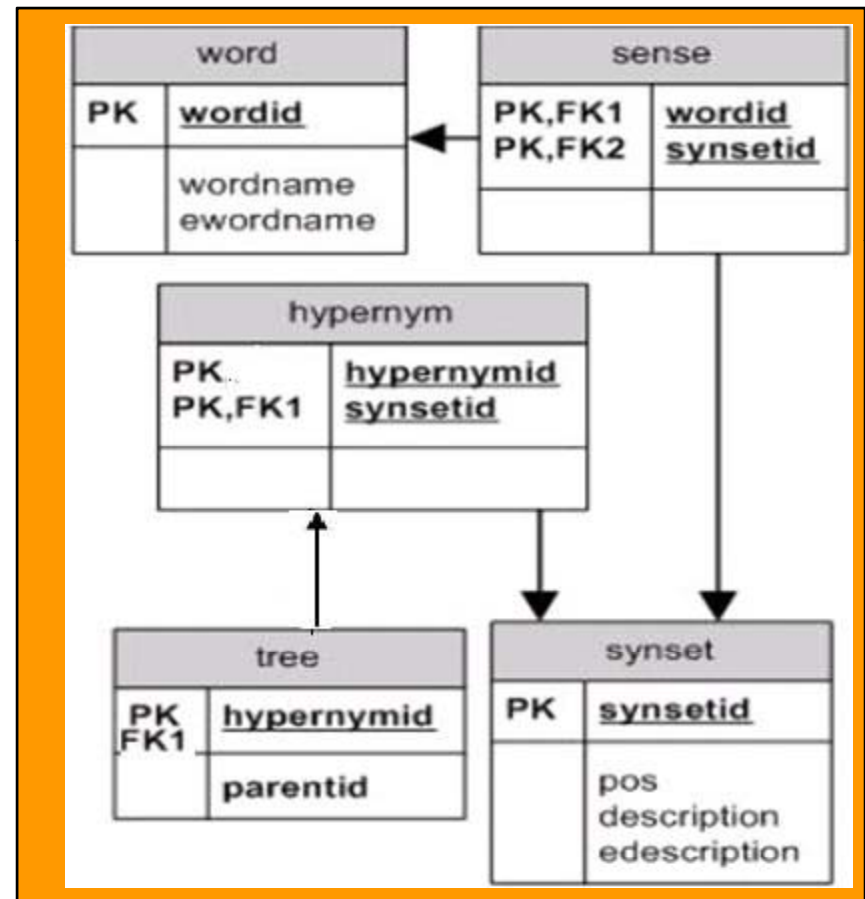
Design and Implementation

- Lexical Source Files
- Grinder
- WordNet Database
- Interface



WordNet Database (WNDB)

- Store the English word to link to other WordNet
- Regenerate the synset at run-time



Interface

- Create the initial database and incrementally update the database
- Use the database to query the data
- Edit the lexical data

Update the database

1st
Interface

Example: “ কাজ।work।কিছু করা বা তৈরির লক্ষে সরাসরি কার্যক্রম।বিশেষ্য।

||work -- (activity directed toward making or doing something)||

hypernyms:। কার্যক্রম | কৃতকর্ম | ঘটিত বিষয় | মনস্তাত্ত্বিক-বিষয় | বিমূর্তন | বিমূর্ত-সত্তা | সত্তা | synonyms:কর্ম, কাজকাম”

1 word table

wordid	wordname	ewordname
1	কাজ	Work
2	কর্ম	Work
3	কাজকাম	Work

3 sense table

wordid	synsetid
1	1
2	1
3	1

4 hypernym table

synsetid	hypernymid
1	2
1	3
1	4
1	5
1	6
1	10
1	8

5 tree table

hypernymid	parentid
2	3
3	4
4	5
5	6
6	10
10	8
8	0

2 synset table

synsetid	description	edescription	Pos
1	কিছু করা বা তৈরির লক্ষে সরাসরি কার্যক্রম	work -- (activity directed toward making or doing something)	বিশেষ্য

Update the database (cont..)

□ BWN's support of incremental update - at this point, a specific complication may arise.

□ Example:

“কালবিন্দু। measure, quantity, amount তাত্ত্বিক সময়।
 বিশেষ্য। ||point, point in time -- (an instant of time)||

hypernyms: |বিমূর্তন | বিমূর্ত-সত্তা | সত্তা | synonyms: কালবিন্দু”

Update the database (cont..)

কালবিন্দু। measure, quantity, amount তাৎক্ষণিক সময়। বিশেষ্য। ||point, point in time
 -- (an instant of time)|| hypernyms: |বিমূর্তন | **বিমূর্ত-সত্তা** | সত্তা | synonyms:
 কালবিন্দু”

1 word table

wordid	wordname	ewordname
12	কালবিন্দু	measure, quantity, amount

2 synset table

synsetid	description	edescription	pos
9	তাৎক্ষণিক সময়	measure, quantity, amount	বিশেষ্য

3 sense table

wordid	synsetid
12	9

4 hypernym table

synsetid	hypernymid
9	-
9	-
9	-

5 tree table

hypernymid	parentid
-	-
-	-
-	-

Update the database (cont..)

- When this “বিমূর্ত-সত্তা” hypernym word comes as an entity
 - Have to update the word and the synset tables with new value, while the sense table remains the same.
- Example:
- অমূর্ত-সত্তা | Abstract Entity | শুধুমাত্র বিমূর্ত (দৈহিক রূপহীন) অস্তিত্ব আছে এমন সত্তা | বিশেষ্য || Abstract Entity -- (an entity that exists only abstractly) || hypernyms: সত্তা | synonyms: অরূপ-সত্তা, নিরূপ-সত্তা, বিমূর্ত-সত্তা, সত্তা |

Update the database (cont..)

1 word table

wordid	wordname	ewordname
12	কালবিন্দু	measure, quantity, amount

2 synset table

synsetid	description	edescription	pos
9	তাৎক্ষণিক সময়	measure, quantity, amount	বিশেষ্য

3 sense table

wordid	synsetid
12	9

1 word table

wordid	wordname	ewordname
13	বিমূর্ত-সত্তা	hypernym

2 synset table

synsetid	description	edescription	pos
10	বিমূর্ত-সত্তা	hypernym	বিশেষ্য

3 sense table

wordid	synsetid
13	10

Update the database (cont..)

1 word table

wordid	wordname	ewordname
1	কাজ	work
2	কর্ম	work
3	কাজকাম	work
7	বিমূর্তন	abstraction
9	সত্তা	entity
12	কালবিন্দু	measure, quantity, amount
13	বিমূর্ত-সত্তা	hypernym

3 sense table

wordid	synsetid
12	9
13	10
7	6
9	8

5 tree table

hypernym mid	parent id
6	10
10	8
8	0

4 hypernym table

synsetid	hypernym mid
9	6
9	10
9	8

Update the database (cont..)

1 word table

wordid	wordname	ewordname
13	বিমূর্ত-সত্তা	Abstract Entity

3 sense table

wordid	synsetid
12	9
13	10

5 tree table

hypernym mid	parent id
6	10
10	8
8	0

2 synset table

synsetid	description	edescription	pos
10	বিমূর্ত-সত্তা	abstract entity -- (an entity that exists only abstractly)	বিশেষ্য

4 hypernym table

synsetid	hypernym mid
9	6
9	10
9	8

Using WNDB

- This the second interface to the WNDB is for querying the data
 - Search option
 - Hypernym relation of a noun

Search Option

2nd
Interface

শব্দ অনুসন্ধান:

অংশ অনুসন্ধান:

বিশেষ্য

There are 2 senses of অংশ

1. অংশ (সম্পর্ক), অবশিষ্ট, বাকি অংশ, শেষ অংশ -- (কোনেকিছুর সাথে সুদূতভাবে সম্পর্কিত কোনো কিছুর অন্তর্ভুক্ত ভাগ।)
2. অংশ (অবস্থান), একাংশ, একপার্শ্ব -- (কোনো কিছুর বর্ধিত বিস্তৃত-অবস্থান।)

Hypernym relation of a noun

2nd Interface

শব্দ অনুসন্ধান:

অংশ অনুসন্ধান:

2 senses of অংশ

Sense 1

অংশ (সম্পর্ক), অবশিষ্ট, বাকি অংশ, শেষ অংশ, -- (কোনোকিছুর সাথে সুদৃঢ়ভাবে সম্পর্কিত কোনো কিছুর অন্তর্ভুক্ত ভাগ।)

=> সম্বন্ধ, সম্পর্ক -- (দুটি সত্তার বা অংশের গুণাবলী দ্বারা সম্পর্কযুক্ত বা অংশভাগী এমন বিমূর্তন সত্তা।)

=> বিমূর্তন -- (সুনির্দিষ্ট উদাহরণাদি থেকে গৃহীত সাধারণ নিদর্শনাদি দ্বারা সৃষ্ট সাধারণ ধারণা।)

=> অমূর্ত-সত্তা, অরূপ-সত্তা, নিরূপ-সত্তা, বিমূর্ত-সত্তা -- (শুধুমাত্র বিমূর্ত (দৈহিক নূপহীন) অস্তিত্ব আছে এমন সত্তা।)

=> সত্তা -- (স্বতন্ত্র অস্তিত্ব আছে এমন ইন্দ্রিয়গ্রাহ্য বা জ্ঞাত বা সিদ্ধান্তকৃত এমন সত্তা কিছু (জীবিত বা জড়)।)

Sense 2

অংশ (অবস্থান), একাংশ, একপার্শ্ব, -- (কোনো কিছুর বর্ধিত বিস্তৃত-অবস্থান।)

=> স্থিতি, অবস্থান -- (কোন সুনির্দিষ্ট পরিসরে অবস্থিত স্থান বা বিস্তৃত স্থান।)

=> দৈহিক লক্ষ্যবস্তু -- (যা স্পর্শ করা যায় এবং দেখা যায় এবং ছায়া প্রদান করে।)

=> দৈহিক সত্তা -- (দৈহিক অস্তিত্ব আছে এমন সত্তা।)

=> সত্তা -- (স্বতন্ত্র অস্তিত্ব আছে এমন ইন্দ্রিয়গ্রাহ্য বা জ্ঞাত বা সিদ্ধান্তকৃত এমন সত্তা কিছু (জীবিত বা জড়)।)

Editing WNDB

3rd
 Interface

BWN

Edit & Delete Operation

Enter Bangla WordName for Edit:

POS:

jLabel7

Bangla WordName:

English WordName:

Description:

English Description:

POS:

Editing WNDB (cont..)

3rd
 Interface

BWN

Edit & Delete Operation

EnterBangla WordName for Edit:

POS:

jLabel7

Bangla WordName:

English WordName:

Description:

English Description:

POS:

Message [X]

The record can't be deleted

Conclusion and Future Work

- Has two key features
 - Incremental building of the WordNet database
 - Editing the WordNet data using an interface
- Extendable to other languages
- Creating Webservice and .NET bindings
- Link to non-Bengali WordNets such as the Hindi and Euro WordNets.

Questions

Thank You