What comes to your mind when I say….. Board?
Similarly what comes to your mind when I say the Sinhala word..... කො?
Understanding the context of the words and their interconnections are important for cognition.

Which helps in the process of drawing conclusions or translating.
The tried and tested way to organize the words in a language to achieve this goal is to create a

**WordNet**
A WordNet for Sinhala

Nisansa de Silva

Advanced Communicator Series – Technical Presentations

Project 1 – The Technical Briefing
What is a WordNet?
WordNet Definition

- A WordNet is a lexical reference system whose design is inspired by current psycholinguistic theories of human lexical memory.
- Nouns, verbs, adverbs and adjectives are organized into synonym sets, each representing one underlying lexical concept.
- Different relations link the synonym sets.
Psycholinguistic Theory

- Human lexical memory for nouns as a hierarchy.
  - Can canary sing? - Pretty fast response.
  - Can canary fly? - Slower response.
  - Does canary have skin? – Slowest response.

Wordnet - a lexical reference system based on psycholinguistic theories of human lexical memory.
Wordnet

- A lexical knowledgebase based on conceptual lookup
  - Organizing concepts in a semantic network.
  - Organize lexical information in terms of word meaning, rather than word form
    - Wordnet can also be used as a thesaurus.
### Wordnet - Lexical Matrix (with examples)

<table>
<thead>
<tr>
<th>Word Meanings</th>
<th>Word Forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>M₁</td>
<td>F₁</td>
</tr>
<tr>
<td></td>
<td>(depend)</td>
</tr>
<tr>
<td></td>
<td>E₁,₁</td>
</tr>
<tr>
<td>M₂</td>
<td>F₂</td>
</tr>
<tr>
<td></td>
<td>(bank)</td>
</tr>
<tr>
<td></td>
<td>E₁,₂</td>
</tr>
<tr>
<td>M₃</td>
<td>F₃</td>
</tr>
<tr>
<td></td>
<td>(rely)</td>
</tr>
<tr>
<td></td>
<td>E₁,₃</td>
</tr>
<tr>
<td>Mₘ</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>Fₙ</td>
</tr>
</tbody>
</table>

**Synonym words**

- (depend) E₁,₁
- (bank) E₁,₂
- (rely) E₁,₃
- (bank) E₂,₂
- (coast) E₂,₃
- (bank) E₃,₂
- (tilt) E₃,₃

**Polysemous Words**
English WordNet developed by Princeton University, U.S.A

- Rich database (large number of synsets)
- Familiarity
- Availability of resources and documentation
- Built on a successful research of human lexical memory
English WordNet developed by Princeton University, U.S.A

Arabic and Japanese WordNets
Language Elements in English WordNet

• Parts of Speech
  • Nouns
  • Verbs
  • Adjectives
  • Adverbs
Language Elements in English WordNet

• Relations
  • Antonymy
  • Hypernym
  • Hyponym
  • Holonymy
  • Meronymy

• Entailment
• Cause
• Similar
• Relational
What is Different in Sinhala?

- **Origin of the Word - Derivation Type**

<table>
<thead>
<tr>
<th>English meaning</th>
<th>thathbawa</th>
<th>thathsama</th>
</tr>
</thead>
<tbody>
<tr>
<td>mother’s</td>
<td>මව්</td>
<td>මාතෘ</td>
</tr>
<tr>
<td>affection</td>
<td>සෙසන/සෙසනසහ</td>
<td>සෙසනසහ</td>
</tr>
<tr>
<td>mother’s affection</td>
<td>මව් සෙසන/සෙසනසහ</td>
<td>මාතෘ ේසේහය</td>
</tr>
</tbody>
</table>

- **Gender**

<table>
<thead>
<tr>
<th>Sinhala sentence</th>
<th>English meaning</th>
<th>Noun used</th>
<th>Verb used</th>
</tr>
</thead>
<tbody>
<tr>
<td>හැංගේ යාවට නැගොසඟ හා</td>
<td>Elder sister got on to the bus</td>
<td>හැංගේ</td>
<td>නැගොසඟ හා (got on)</td>
</tr>
<tr>
<td>අයියේ යාවට නැගොසඟ හා</td>
<td>Elder brother got on to the bus</td>
<td>අයියේ</td>
<td>නැගොසඟ හා (got on)</td>
</tr>
</tbody>
</table>
What is Different in Sinhala?

**Written Spoken Difference**

<table>
<thead>
<tr>
<th>English</th>
<th>Sinhala spoken</th>
<th>Sinhala written</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eyes</td>
<td>ආයේ දැමි</td>
<td>අරදාමූ, කොහෝලෑ</td>
</tr>
<tr>
<td>The sun</td>
<td>අරු</td>
<td>අරු</td>
</tr>
</tbody>
</table>

**Word Root**

<table>
<thead>
<tr>
<th>Sinhala</th>
<th>Sinhala</th>
<th>the man</th>
<th>Sinhala</th>
<th>the men</th>
</tr>
</thead>
<tbody>
<tr>
<td>අඩිඩ්හලී</td>
<td>අඩිඩ්හලී</td>
<td>a man</td>
<td>අඩිඩ්හලී</td>
<td>men</td>
</tr>
<tr>
<td>අඩිඩ්හලී ශාලීයයන්</td>
<td>අඩිඩ්හලී ශාලීයන්</td>
<td>by the man</td>
<td>අඩිඩ්හලී ශාලීයන්</td>
<td>by the men</td>
</tr>
<tr>
<td>අඩිඩ්හලී විසියට</td>
<td>අඩිඩ්හලී විසියට</td>
<td>to the man</td>
<td>අඩිඩ්හලී විසියට</td>
<td>to the men</td>
</tr>
<tr>
<td>අඩිඩ්හලී විසියණාවය</td>
<td>අඩිඩ්හලී විසියණාවය</td>
<td>from the man</td>
<td>අඩිඩ්හලී විසියණාවය</td>
<td>from the men</td>
</tr>
<tr>
<td>අඩිඩ්හලී විසියණාවය</td>
<td>අඩිඩ්හලී විසියණාවය</td>
<td>the man’s</td>
<td>අඩිඩ්හලී විසියණාවය</td>
<td>the men’s</td>
</tr>
<tr>
<td>අඩිඩ්හලී</td>
<td>අඩිඩ්හලී</td>
<td>man</td>
<td>අඩිඩ්හලී</td>
<td>men</td>
</tr>
</tbody>
</table>
Approach
WordNet Constructor Architecture

Presentation Layer
- Crowdsourcing Interface
- Sinhala WordNet Browser

Process Layer
- WordNet Constructor Core
- MongoDB to Text DB Transformer
- Sinhala WordNet API

Data Access Layer
- API for Linguistic Resources
- English WordNet API
- MongoDB Data Access API

Linguistic Resources
- English WordNet Text Database
- Sinhala WordNet Text Database
WordNet Constructor

A crowdsourcing WordNet development platform designed for Sinhala language which can be extended to develop a WordNet for any other language with minor configuration.
Crowdsourcing Constructor Core

• A façade to English WordNet API
• Facilitating Sinhala specific features
• Expanding English WordNet
Resource API

• Proving suggestions
• Madhura dictionary
  • Merging the results of the set of words in a synset

• Sinhala Concise of Sinhala Dictionary Office, Sri Lanka
Sinhala WordNet API

• Retrieve Sinhala synsets and support search in the Sinhala WordNet browsers for Sinhala WordNet end users.

• Provides a rich API for Sinhala Natural Language Processing developers to use the Sinhala WordNet.
User Interface
Target Users

• Non-English WordNet developers
• General Sinhala users
• Sinhala Linguists
• Sinhala NLP (natural Language Processing) developers
Milestones

• Long paper accepted at 7th Global WordNet Association Conference (in Tartu, Estonia, January 25-29, 2014)
• Mini ERU – Poster Presentation
• NBQSA- Participation
• SIIIP- Participation
Project Deliverables

• A generic system for developing a WordNet for a given language expanding English WordNet
• The Sinhala WordNet
• Java API for the Sinhala WordNet
• Desktop Browser Application for the Sinhala WordNet
• Online Browser for the Sinhala WordNet
Group

• Members
  • Malaka Gallage
  • Indeewari Wijesiri
  • Buddhika Gunathilaka
  • Madhuranga Lakjeewa

• Technical Supervisors
  • Dr. Daya Chinthana WImalasuriya
  • Prof. Gihan Dias
  • Mr. Nisansa Dilushan de Silva
Q & A