Pali lexicography using lexonomy

Anupama Devendra Sakhare¹, Arshad Shareef² and Varsha Rangari³

¹Assistant Professor, Department of Electronics and Computer Science, RTMNU, Nagpur
²Senior Faculty- IT, Oberoi Centre for Excellence (Cloud Technology), RTMNU, Nagpur
³Associate Professor, Department of Electronics, M P Deo Memorial Science College, Nagpur

Email: adsakhare616@gmail.com

Abstract. Lexonomy is a platform which is an open source for writing and publishing dictionaries. Pali Lexicography is actually a dictionary created by using Lexonomy, which is a web based tool for creating dictionaries. Since a dictionary is a means of understanding any language and its translation, the present work is about Pali Lexicography. It is a project that has been created using simple XML tags and technical skills without any cost. In this, word mapping is done using selected words that are chosen from the set of Pali Language which are also present in the Marathi Language set having almost similar meaning in Marathi Language. All the Pali words have their English meaning which are automatically sorted generating the Lexicographic order. Pali dictionary is created with a simple user interface and is easily operable.

1. Introduction
Lexonomy is an open source tool which is available for free facilitating to write and publish various dictionaries. It can be operated using a web browser where data storage is on the server rather than on a standalone computer. With the increase of technology across the globe, there are instances wherein a group of members working on some international project need to communicate using an interface that is understandable by everyone. For this, an interlinguistic tool can do the job. As of now, there are many monolingual and bilingual tools available but due to intervention of computational linguistics there is an essence for an interlinguistic tool. Keeping in view of the existing on-line digital dictionaries that can also solve the purpose, there is also a need to link these dictionaries to the lexicographic order that can certainly benefit the user in many ways.

For implementing this, we have taken the corpus from Pali language, some of the words which are also in the Marathi language having the similar meaning. Then mapping is done using the help of Pali words which have their English language meaning too and gets sorted automatically using the language constructs of translation. A user can view this with the help of a tool creating using XML and operate via a user interface.

1.1. Subentries
Lexonomy provides the option of subentries. That is, during the definition of entry schema, we can provide labels to few sections of the XML tree, for example we can make phraseological subentries, to be ‘shareable’, which can be shown into snippets of XML thus allowing it to get displayed in multiple entries at the same time. All the elements listed in this context shall be subentries by Lexonomy, and are shareable among several entries. Shareable subentries have a shaded background in the editing interface and are followed by a button which tells number of other entries, apart from the existing ones.

1.2. Headwords
This part allows to configure the headword list and its display. You can select the element which contains the entry's headword. If no selection is made then Lexonomy will be able to guess what the headword of each entry is. You can also adjust the order of the headwords in the entry list by specifying the element that should be used for sorting, e.g. part of speech and you can hard-code the alphabetical order that should be used for sorting.
One can also specify any elements whose content you want be displayed beside the headword in the entry list, such as homograph numbers or part-of-speech labels.

1.3. Flags
We can set different flags (coloured circles) to each entry in order to be able to easily monitor its status. This can then be used to control the workflow in the project. To be able to use this function, each entry needs to have a specific element containing one of the values you have selected for your workflow. For example, let's say you have three stages:

1. automatic entries not yet checked by lexicographers: status = AE (automatic entry)
2. entries checked by a lexicographer: status = L (lexicographer)
3. entries confirmed by the editor: status = C (complete)

2. Literature Survey
2.1. Trenckner’s – A Critical Pali Dictionary
In the year 1855, the first Pali text scientific edition full volume has been released in the form of transcription with Latin character and is widely accepted Pali publication. Apart from this, Trenckner left two volumes containing publishing work for a complete Pāli Grammar namely morphology, phonology, and prosody, but this, as well as his lexicon and the edition of Majjhima-Nikāya and its commentary, was unfortunately interrupted due to his death. His valuable library was sold to Germany, all his transcripts were submitted to some University Library, while his lexicon-papers was generously given by his executor testamenti to the writer, who after sorting in order, had them deposited in the Library with the other manuscripts.

2.2. Pali Text Society – Journal of the Pali Text Society
In this study, the true lexicographical text, deals with each word in a way which supplies the place of a dictionary. Practically every word is clarified, either by synonyms, or by a statement of word’s grammatical form and function. A comparable practice is followed in different sections of this journal, and in all these texts, the primary aim is to clarify the words of the Buddha, to specify their acceptation and then to specify those words with overall teaching, not only to define what it means but also what it means in the context of Dhamma as a whole.

2.3. A Pali Reader with Notes – Muang Tin
This publication contains the selections from Jatakas, i.e. Buddhist Birth Stories and Mangalasutta i.e. sermon on blessings in common with the old course. The new chapters included in this study are: Lokaniti which means worldly behaviour, Kapirajacariya Sasapanditacariya which is the conduct of the monkey-king and wise hare, Metta Sutta which implies sermon on love and Singalovada Suttanta which states sermon on advice to singala. Among these, Lokaniti is popular in Burma and has been included at the wish of Burman Buddhists.

3. Technical Implementation
3.1. Method for Dictionary Creation
The stepwise procedure is illustrated as follows:
1. We use xml script for mentioning headword along with its translation and the source of origin and finally specifying the target.
2. All these operations require an entry ID which is mentioned at the start of the processing.
3. Multiple words are given as input for the translation to be done as mentioned in Step-1.
4. The tool then generates the corresponding dictionary meaning in English required for understanding the given Pali word.
5. Sorting of words in lexicographical order is also done at the last operation.

- `<entry xmlns:lxnm="url" lxnm:entryID="607">
  <headword xml:space="preserve">Parami</headword>
  <translation xml:space="preserve">Spiritual perfections</translation>
  `<example>`
  `<source>`
  `[Ten parami]`
  `<target>`
  `</target>`
  `</example>`
  `</sense>`
  `</entry>`

- `<entry xmlns:lxnm="url" lxnm:entryID="609">
  <headword xml:space="preserve">Abhassara</headword>
  `<source>`
  `[The Radiant One]`
  `<target>`
  `</target>`
  `</entry>`

- `<entry xmlns:lxnm="url" lxnm:entryID="610">
  <headword xml:space="preserve">Brahmacariya</headword>
  `<source>`
  `[Pure Chaste]`
  `<target>`
  `</target>`
  `</entry>`

- `<entry xmlns:lxnm="url" lxnm:entryID="611">
  <headword xml:space="preserve">Citta</headword>
  `<source>`
  `[Mind]`
  `<target>`
  `</target>`
  `</entry>`

- `<entry xmlns:lxnm="url" lxnm:entryID="612">
  <headword xml:space="preserve">Deva Dutta</headword>
  `<source>`
  `[Divine Messengers]`
  `<target>`
  `</target>`
  `</entry>`

- `<entry xmlns:lxnm="url" lxnm:entryID="613">
  <headword xml:space="preserve">Eka</headword>
  `<source>`
  `[One]`
  `<target>`
  `</target>`
  `</entry>`
- <entry xmlns:lxnm="url" lxnm:entryID="614">
  <headword xml:space="preserve">Gati</headword>
  - <sense>
    <translation xml:space="preserve">Existence</translation>
    - <example>
      <source xml:space="preserve">Sugati</source>
      <target xml:space="preserve">Duggati</target>
    </example>
  </sense>
</entry>

- <entry xmlns:lxnm="url" lxnm:entryID="615">
  <headword xml:space="preserve">Hetu</headword>
  - <sense>
    <translation xml:space="preserve">Cause</translation>
  </sense>
</entry>

- <entry xmlns:lxnm="url" lxnm:entryID="616">
  <headword xml:space="preserve">Indriya</headword>
  <sense>
    <translation xml:space="preserve">Sense Organs</translation>
  </sense>
</entry>

- <entry xmlns:lxnm="url" lxnm:entryID="617">
  <headword xml:space="preserve">Jara</headword>
  - <sense>
    <translation xml:space="preserve">Old age</translation>
  </sense>
</entry>

- <entry xmlns:lxnm="url" lxnm:entryID="618">
  <headword xml:space="preserve">Kamma</headword>
  - <sense>
    <translation xml:space="preserve">Action</translation>
  </sense>
</entry>

- <entry xmlns:lxnm="url" lxnm:entryID="619">
  <headword xml:space="preserve">Kusala</headword>
  - <sense>
    <translation xml:space="preserve">Morally Good</translation>
  </sense>
</entry>

- <entry xmlns:lxnm="url" lxnm:entryID="620">
  <headword xml:space="preserve">Lobha</headword>
  - <sense>
    <translation xml:space="preserve">Greed</translation>
  </sense>
</entry>

- <entry xmlns:lxnm="url" lxnm:entryID="621">
  <headword xml:space="preserve">Mangala</headword>
  - <sense>
    <translation xml:space="preserve">Auspicious</translation>
  </sense>
</entry>
- <entry xmlns:lxnm="url" lxnm:entryID="622">
  <headword xml:space="preserve">Nibbana</headword>
  - <sense>
    <translation xml:space="preserve">Nirvana</translation>
  </sense>
</entry>

- <entry xmlns:lxnm="url" lxnm:entryID="623">
  <headword xml:space="preserve">Punya</headword>
  - <sense>
    <translation xml:space="preserve">To do Good Merit</translation>
  </sense>
</entry>

- <entry xmlns:lxnm="url" lxnm:entryID="624">
  <headword xml:space="preserve">Rupa</headword>
  - <sense>
    <translation xml:space="preserve">Visual Object</translation>
  </sense>
</entry>

- <entry xmlns:lxnm="url" lxnm:entryID="625">
  <headword xml:space="preserve">Sacca</headword>
  - <sense>
    <translation xml:space="preserve">Truth</translation>
  </sense>
</entry>

- <entry xmlns:lxnm="url" lxnm:entryID="626">
  <headword xml:space="preserve">Tipitaka</headword>
  - <sense>
    <translation xml:space="preserve">3 Baskets</translation>
  </sense>
</entry>

- <entry xmlns:lxnm="url" lxnm:entryID="627">
  <headword xml:space="preserve">Uposatha</headword>
  - <sense>
    <translation xml:space="preserve">Fasting</translation>
  </sense>
</entry>

url: http://www.lexonomy.eu/

**Figure 3.1.** Code used for creating the Dictionary

### 3.2. Sketch Engine Connection

In this section you will find options for configuring your connection with the Sketch Engine. Sketch Engine URL, Sketch Engine API URL and Corpus name are mandatory fields, others are optional.
3.3. Examples
Here you configure the pulling of example sentences from Sketch Engine. You need to select the element in which every example will be wrapped, then provide an XML template for every example that will be inserted into your entries, and select the element that marks the headword in the entries, if relevant.

4. Results

Figure 4.1. Words used for creating the Dictionary

Figure 4.2. XML tags used for creating the Dictionary
5. Conclusion

Pali Lexicography is a dictionary created by using Lexonomy, which is a Dictionary writing system. The present work is about Pali Lexicography – dictionary project that has been created using simple XML editing tags and simple technical skills with no financial cost. Apart from meaning translation, sorting of the words in lexicographical order is also done.
6. References

[1] Mechura Michal, Introducing Lexonomy: an open-source dictionary writing and publishing system, proceedings of elex 2017 Conference, Lexical Computing 2017, p. 662-679, 18 pp ISSN2533-5626
[2] https://cpd.uni-koeln.de/intro/trenckner [p.III]
[3] Pali Text Society – Journal of the Pali Text Society, Vol. XXII, p.06
[4] A Pali Reader with Notes – Muang Tin, 2nd Edition, 1920
[5] Web Resources: http://www.elexicography.eu/
[6] https://opensuse.org/licenses/MIT
[7] https://github.com/michmech/Lexonomy
[8] https://nodejs.org/
[9] https://www.Sketch engine co. uk