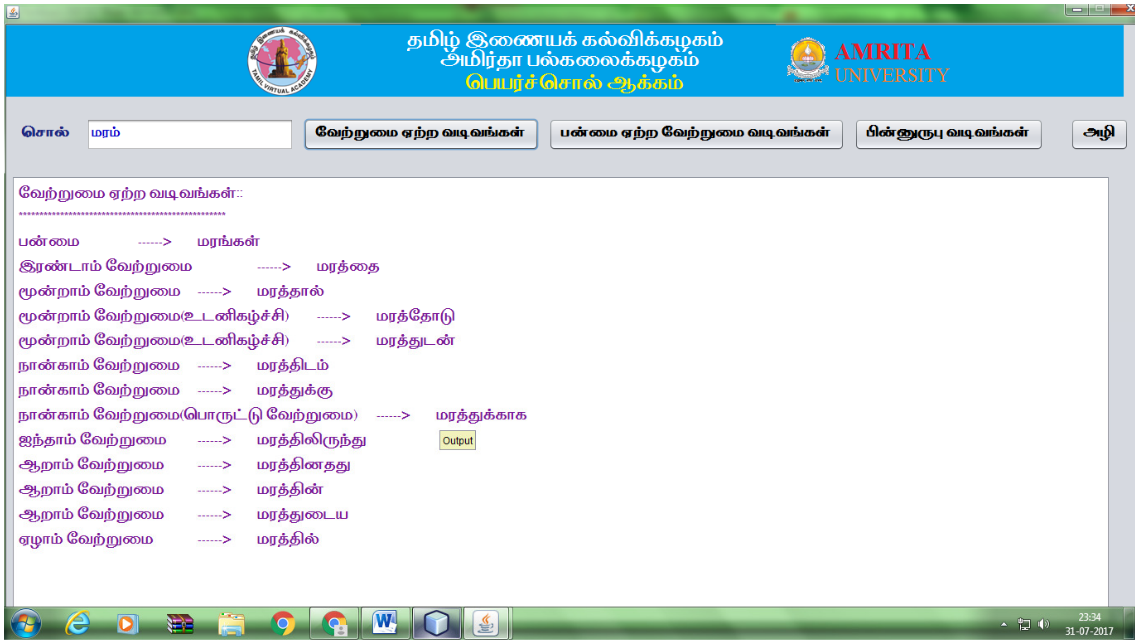
Final phase Report of the project on “Computing Tools for Tamil Language teaching and learning”

Dr. Anand Kumar M, Rajendran S, Soman K P

CEN, Amrita Vishwa Vidyapeetham, Coimbatore

I. Tool for learning Verb Conjugation and Noun Declension

A Tool for learning Verb Conjugation and Noun Declension is getting developed as one of the components of the present project. Tamil is a morphologically rich language. Being agglutinative language most of the grammatical information are expressed by suffixes. For example, nouns are inflected for number and cases and verb are inflected for tenses, moods and aspects and subject agreement markers. A morphological generator capturing conjugation of verbs have been developed as one of components of the present project. The Morphological Generator takes lemma and grammatical information as input and gives inflected forms of the given word. It is a reverse process of Morphological Analyzer. Morphological generator system implemented here is a rule based system which makes use of morpheme concatenating rules and gives us the all the conjugated forms of a given verb and declension forms of a given noun. In the first phase only "verb conjugation" has been completed. We will take up noun declension in the second phase. We hope to develop a full-fledged tool which helps Tamil learners in understanding verb conjugation and noun declension in Tamil.

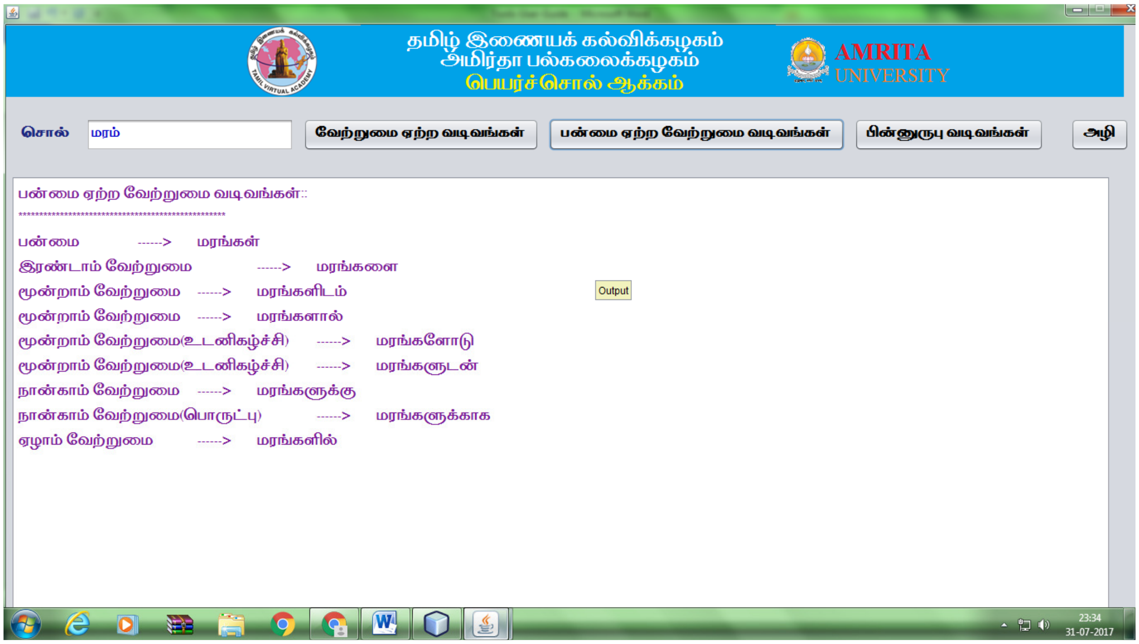


சொல் மரம்

```
வேற்றுமை ஏற்ற வடிவங்கள்::
****************************************************

பன்மை          ------>    மரங்கள்
இரண்டாம் வேற்றுமை          ------>    மரத்தை
மூன்றாம் வேற்றுமை    ------>    மரத்தால்
மூன்றாம் வேற்றுமை(உடனிகழ்ச்சி)    ------>    மரத்தோடு
மூன்றாம் வேற்றுமை(உடனிகழ்ச்சி)    ------>    மரத்துடன்
நான்காம் வேற்றுமை    ------>    மரத்திடம்
நான்காம் வேற்றுமை    ------>    மரத்துக்கு
நான்காம் வேற்றுமை(பொருட்டு வேற்றுமை)    ------>    மரத்துக்காக
ஐந்தாம் வேற்றுமை    ------>    மரத்திலிருந்து
ஆறாம் வேற்றுமை    ------>    மரத்தினதது
ஆறாம் வேற்றுமை    ------>    மரத்தின்
ஆறாம் வேற்றுமை    ------>    மரத்துடைய
ஏழாம் வேற்றுமை    ------>    மரத்தில்
```

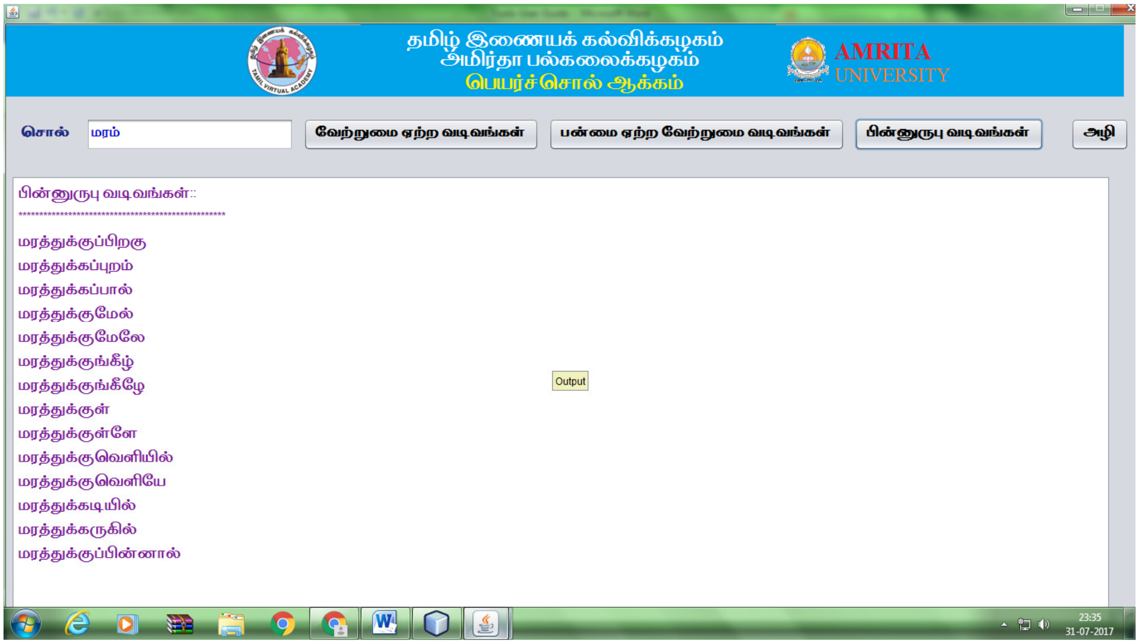
Fig 1.1பெயர்ச்சொல்ஆக்கம்- வேற்றுமைஏற்ற வடிவங்கள்



```
பன்மை ஏற்ற வேற்றுமை வடிவங்கள்::
**************************************************

பன்மை          ------>    மரங்கள்
இரண்டாம் வேற்றுமை          ------>    மரங்களை
மூன்றாம் வேற்றுமை    ------>    மரங்களிடம்
மூன்றாம் வேற்றுமை    ------>    மரங்களால்
மூன்றாம் வேற்றுமை(உடனிகழ்ச்சி)    ------>    மரங்களோடு
மூன்றாம் வேற்றுமை(உடனிகழ்ச்சி)    ------>    மரங்களுடன்
நான்காம் வேற்றுமை    ------>    மரங்களுக்கு
நான்காம் வேற்றுமை(பொருட்பு)        ------>    மரங்களுக்காக
ஏழாம் வேற்றுமை       ------>    மரங்களில்
```

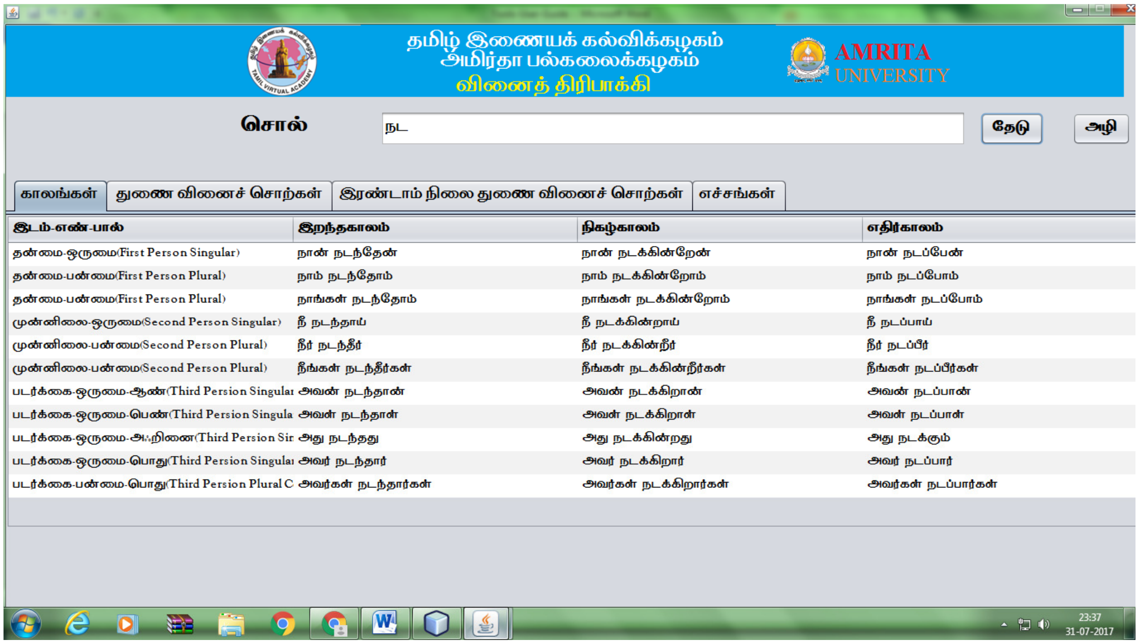
Fig 1.2பெயர்ச்சொல்ஆக்கம்- பன்மைஏற்றவேற்றுமை வடிவங்கள்



```
பின்னுருபு வடிவங்கள்::
**************************************************

மரத்துக்குப்பிறகு
மரத்துக்கப்புறம்
மரத்துக்கப்பால்
மரத்துக்குமேல்
மரத்துக்குமேலே
மரத்துக்குங்கீழ்
மரத்துக்குங்கீழே
மரத்துக்குள்
மரத்துக்குள்ளே
மரத்துக்குவெளியில்
மரத்துக்குவெளியே
மரத்துக்கடியில்
மரத்துக்கருகில்
மரத்துக்குப்பின்னால்
```

Fig 1.3பெயர்ச்சொல்ஆக்கம்- பின்னுருபு வடிவங்கள்



சொல் நட

இடம்-எண்-பால்	இறந்தகாலம்	நிகழ்காலம்	எதிர்காலம்
தன்மை-ஒருமை(First Person Singular)	நான் நடந்தேன்	நான் நடக்கின்றேன்	நான் நடப்பேன்
தன்மை-பன்மை(First Person Plural)	நாம் நடந்தோம்	நாம் நடக்கின்றோம்	நாம் நடப்போம்
தன்மை-பன்மை(First Person Plural)	நாங்கள் நடந்தோம்	நாங்கள் நடக்கின்றோம்	நாங்கள் நடப்போம்
முன்னிலை-ஒருமை(Second Person Singular)	நீ நடந்தாய்	நீ நடக்கின்றாய்	நீ நடப்பாய்
முன்னிலை-பன்மை(Second Person Plural)	நீர் நடந்தீர்	நீர் நடக்கின்றீர்	நீர் நடப்பீர்
முன்னிலை-பன்மை(Second Person Plural)	நீங்கள் நடந்தீர்கள்	நீங்கள் நடக்கின்றீர்கள்	நீங்கள் நடப்பீர்கள்
படர்க்கை-ஒருமை-ஆண்(Third Persion Singular	அவன் நடந்தான்	அவன் நடக்கிறான்	அவன் நடப்பான்
படர்க்கை-ஒருமை-பெண்(Third Persion Singula	அவள் நடந்தாள்	அவள் நடக்கிறாள்	அவள் நடப்பாள்
படர்க்கை-ஒருமை-அஃறிணை(Third Persion Sin	அது நடந்தது	அது நடக்கின்றது	அது நடக்கும்
படர்க்கை-ஒருமை-பொது(Third Persion Singular	அவர் நடந்தார்	அவர் நடக்கிறார்	அவர் நடப்பார்
படர்க்கை-பன்மை-பொது(Third Persion Plural C	அவர்கள் நடந்தார்கள்	அவர்கள் நடக்கிறார்கள்	அவர்கள் நடப்பார்கள்

Fig 1.4வினைத்திரிபாக்கி

II. Tool for developing vocabulary skill using Ontology of Tamil vocabulary

Tamil Onto-thesaurus is an outcome of a very long research activity went on in the field of lexical semantics of Tamil vocabulary. It went through several stages before being culminated into Tamil onto-thesaurus. It depicts our travel from Tamil thesaurus to Tamil word net. It is a lexical resource which amalgamates all sorts of information available in a dictionary, thesaurus and word net.A paper thesaurus for Tamil was prepared in 1990 based on the principles of componential analysis of meaning propounded by Nida (1975) and was published in 2001 (Rajendran, 2001), nearly after a decade. Following the paper thesaurus, an Electronic thesaurus for Tamil was attempted and a book on Tamil electronic thesaurus was published in 2006 (Rajendran and Baskaran, 2006) The preparation of wordNet for Tamil was undertaken (2001-2003) with the financial assistance from Tamil Virtual University (renamed now as Tamil virtual academy) and a crude version of it based on the ontology developed by Rajendran (Rajendran, 2001) was submitted to the institute in 2003. After that, from 2009 onwards with the fund received from MHRD and Department of electronics and information Technology of Govt. of India the building of Dravidian wordNet was executed based on Hindi wordNet; nearly 3000 synsets (concepts) have been completed. Still we have to go a long way to achieve the desired target. At present a team from CEN, Amrita University is involved in building onto-thesaurus for Tamil as a part of the project entitled “Computing Tools for Tamil Language teaching and learning”.The project is funded by Tamil Virtual Academy, Chennai.

Tamil Onto-thesaurus

Thesaurus is a in its wider sense is a classification of words by concepts, topics, or subjects. The present Tamil Onto-thesaurus is the extended version of Electronic thesaurus of Tamil focusing more on the ontological features. Two kinds of issues arise in the preparation of Tamil onto-thesaurus:

- Linguistic issues
- Computational issues

Linguistic Issues

It involves mainly the following four tasks:

1. Developing an ontology for Tamil based on structural semantic principles.
2. Establishing semantic domains and sub domains based on distinguishing semantic or Componential features of lexical items.
3. Classifying Tamil vocabulary to fit into the ontology developed.
4. Linking words by various semantic or lexical relations such as synonymy, hyponymy-hyperonymy, meronymy-holonymy, compatibility, and incompatibiliity.

Computational Issues

It involves mainly the following three tasks:

1. Conversion of linguistic data base into computer accessible format.
2. Preparation of a tool to provide the facilities for augmenting, entering and editing the raw data, and classifying the lexical items in a semi-automatic way.
3. Creation of user friendly interfaces for accessing the onto-thesaurus in simple manner.

Ontology of Tamil vocabulary

The ontology available in Rajendran (2001), which is founded on the theory of componential analysis of meaning propounded by Nida (1975) is enhanced to suit the present purpose. The following is the skeletal structure of the Tamil ontology adopted in Onto-thesaurus.

Ontology Relationships

Relationship between nodes in the ontology tree has been classified in to two types namely IS-A relationship and
HAS-A relationship.
Both of the relationships are frequently used in hierarchical to show the link between the nodes.

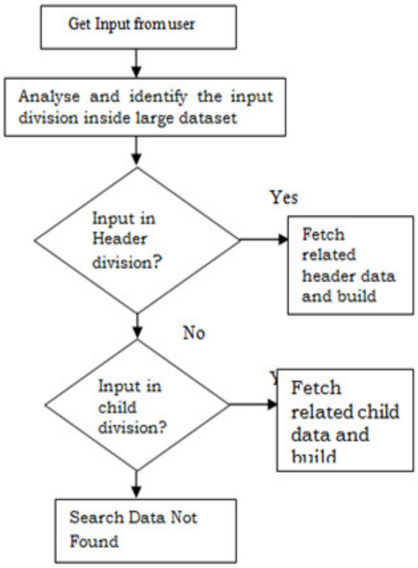
- *IS-A Relation*
  This types specifies the relationship between parent nodes and child nodes in the ontology tree. *Eg. Crow **is a** Bird, Dog **is a** Animal*. Here using this is-a relationship, one can easily identify Crow is a child node of Bird node and Dog is a child node of Animal node.

- *HAS-A Relation*
  It shows the properties of a node or entity. Each node has its own properties and it can be expressed with has-a relationship notations. E.g.: Crow has wings.

- *Sibling Relation*
  When a node entity A has a direct child B, then B is sibling for the entity A. i.e. Sibling refers t a direct child for a node.

- *Transitive Relation*
  Transitivity between the nodes also possible in the ontology tree. It comes like a node A->B->C, So
  A-->C. This type of relations is called transitive.

In the above hierarchy, there exist both transitive and sibilings relation. The node விண்வெளி has three sibilings namely (வான்வெளி, ஆகாயவெளி, வானம்), since all these three nodes are direct child. There is not sibilings exist for the nodes (வான்வெளி, ஆகாயவெளி, வானம்). Transitive relations comes between the nodes விண்வெளி ->வானம் ->ஆகாயம்,வின்,வான். Node entity விண்வெளி has a direct child as வானம் and வானம் has a three direct child ஆகாயம்,வின்,வான். So using transitive rule, it can be structured as விண்வெளி - >ஆகாயம்வின்வான்.


Get Input from user
Analyse and identify the input division inside large dataset
Input in Header division?
Yes
Fetch related header data and build
No
Input in child division?
Y
Fetch related child data and build
Search Data Not Found

Building Onltology Framework

Generating ontology tree from a given huge data corpus based on the user query is a backbone activities for this intelligent information retrieval and visualization process. This type of tool provides a powerful way for data representation and knowledge mining process. User can input search word to the system, then the system will analyze the dataset based on the word and identify the location of the given input word inside the dataset. If the input is located in the header division, then the system fetch the related header node, else if the input is located in the non-header division then the related data from the non-header division is fetched to generate the simple hierarchy and ontology tree. The following are the flow steps and proposed algorithm to build ontology and to visualize it effectively.

The following are the data representation label used in the ontology tree generation process.

Notations	Tamil Equivalent	Symbols
Synonym	இணைமொழியம்	<>
Hyponym	வகைமொழியம்	{ }
Meronym	பகுதிமொழியம்	/ |
Gender	பால்மொழியம்	$#
Antonym	எதிர்மொழியம்	*
Derived	பெயராக்கமொழியம்	~ `
Adv	வினையுரிச்சொல்	^ !

- *Synonym<>*

  It represents the different words which gives same meaning. It shows one to many relationship in the hierarchy. Eg:Sysnonym<>சூரியன், <ஞாயிறு, ஆதவன், பகலவன், திவாகரன், ஆதித்தன், கதிரவன், கதிரோன், கிரணன், செங்கதிரோன், செங்கதிர், வெங்கதிரோன், வெங்கதிர், வெஞ்சுடர்ஆயிரங்கதிரோன், ஆயிரங்கிரணன், உலகநேந்திரன், உலகப்பாந்தவன்>; { எழுஞாயிறு, ஏறுஞாயிறு; சாய்ஞாயிறு, இறங்குஞாயிறு; இளஞ்சூரியன், <பாலசூரியன்>; }.

- *Hyponym* { }

  சப்தமண்டலம்2: { வாயுமண்டலம்; ருணமண்டலம்; சந்திரமண்டலம்; சூரியமண்டலம்; ட்சத்திரமண்டலம்; அக்கினிமண்டலம்; திரிசங்குமண்டலம்; }

- *Meronym* / |

  காற்றுமண்டலம்:/வெளிமண்டலம்;வெப்பமண்டலம், உஷ்ணமண்டலம்;படலம்:அயனிப்படலம்;ஓசோன்படலம்; |

- *Gender* $ #

  இளைஞர்கள்: {இளைஞர் $ இளைஞன் #, வாலிபர் $ வாலிபன் #, வாலிபப்பையன், இளவல், இளந்தாரி, இளைஞோர் $ இளைஞோன் #, காளை, விடலை, இளவட்டம், வயசுப்பையன்,வயசுப்பிள்ளை, வயசுப்பிள்ளையாண்டன், இளம்வயதினன்-இளம்வயதினர், இளம்பருவத்தார் $ இளம்பருவத்தான்#, இளம்பிராயத்தார் $ இளம்பியாயத்தான்#, பதின்பருவத்தினர் $ பதின்பருவத்தினன் #, பதின்வயதினர் $ பதின்வயதினன் #, வயதுவந்தோர், <வயதுவந்தவர்> $ வயதுவந்தவன் #, இளந்தை, இளந்தாரி, வாலியன், பாலியன்; இளரத்தம்> }

- *Derived* ~ `

நிகழ்வுகள்குறித்தவை: { நிகழ்; ~ நிகழ்தல்; நிகழ்ச்சி; நிகழ்வு; ` <சம்பவி; ~ சம்பவித்தல்; சம்பவம்; ` விளை; > ~ விளைதல்; விளைவு; ` >நேர்; ~ நேர்தல்; ` பலி; ~ பலித்தல்; ` ஈடேறு; ~ ஈடேறுதல்; ஈடேற்றம்; ` நிலவு; ~ நிலவுதல்; ` }

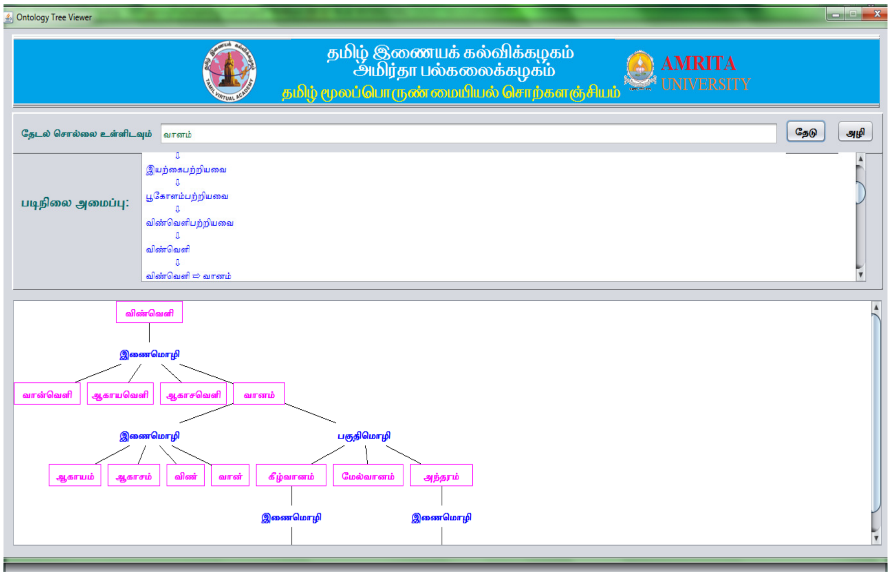
- *Antonym% **
  It gives you opposite meaning for a word. It shows one to one relationship in the hierarchy. *Eg:ஆண்மரம் - >பெண்மரம்.*

- *Adv/Adj^ !*
  நீண்டகாலம்தொடர்பானவை: { நீண்டகாலம்; ^ நீண்டகாலமாக !: நீடூழி, <நெடுங்காலம் ^ நெடுங்காலமாக; ! ரெம்பகாலம்; ^ ரெம்பகாலமாக; ! நீண்டநேரம், <நெடுநேரம், ரெம்பநேரம், நீண்டநேரம், நிறையநேரம்; >நீண்டநாள், <நெடுநாள்; ரெம்பநாள், நிறையநாள்>; பல்லாண்டு;நூறாண்டு, <நூற்றாண்டு> }

Onto-thesaurus is a knowledge representation and these knowledge are visualized in the form of clusters instead of hierarchical tree. Each parent entity and its all available sibling entities re-grouped on to a cluster and all clusters generated during search patterns are linked with the label entities.

CONCLUSION

Presently we developed a real time system for Ontology based intelligent system for information retrieval. This system works with huge datasets which contains 50,000 word count for more general domain. So it a word based information retrieval which gives hierarchy and ontology tree for the user. We don't have any mathematical related computational works right now in this system and it can be accommodated in future research works. Based on requirement, the same system can also be used for domain based dataset and any other use cases. Since the implementation has been done generic manner which support any languages. Onto-thesaurus can be a very efficient tool for visualizing the dataset from large data corpus. There data corpus are represented using XML format.In future, such implementation for ontology based intelligent system can be integrated with probability graphical model (PGM) to improve the existing features with mathematical modeling to shows the relationship between the nodes in the hierarchy based on probability distributions. Apart from that, distance between the node can also be calculated from the generated ontology tree. These information's can be pretty much useful for further ontology research and improve more knowledge mining during representation.



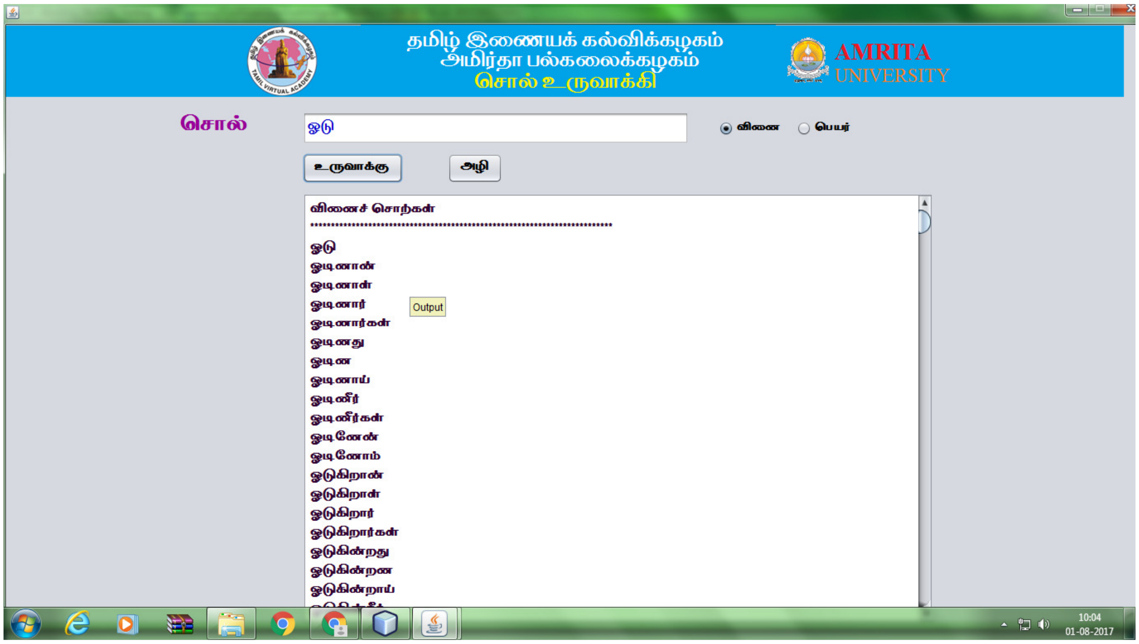
தேடல் சொல்லை உள்ளிடவும் வானம்

படிநிலை அமைப்பு:

```
⇩
இயற்கைபற்றியவை
⇩
பூகோளம்பற்றியவை
⇩
விண்வெளிபற்றியவை
⇩
விண்வெளி
⇩
விண்வெளி ⇒ வானம்
```

Fig 2.1மூலப்பொருண்மையியல்

III. Tool for learning Morphology generator



சொல் உருவாக்கி

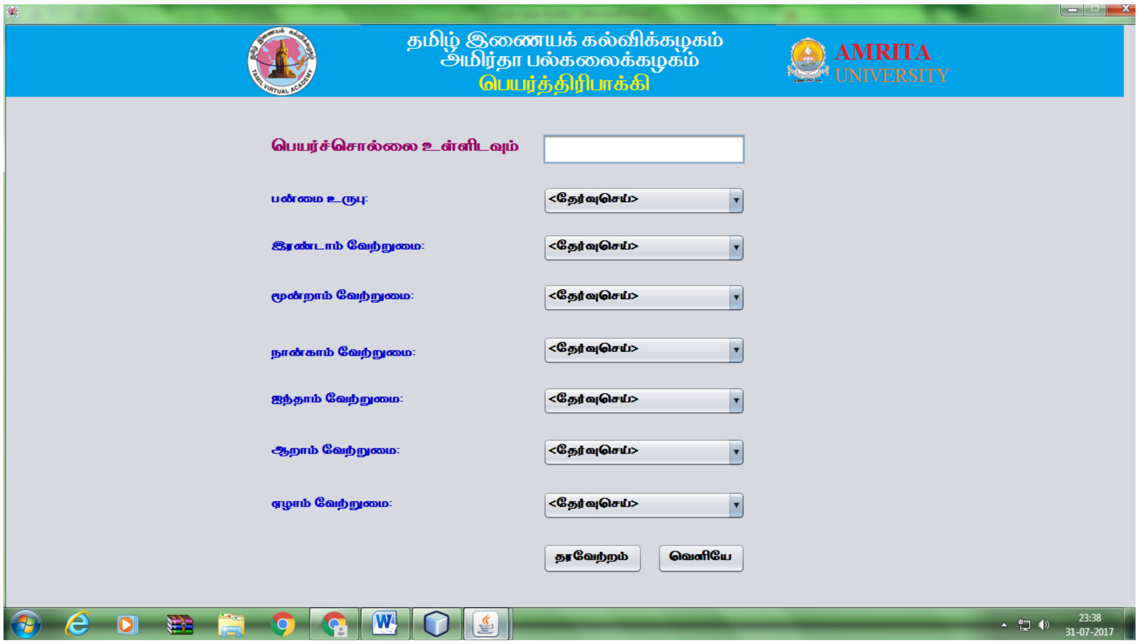
சொல்: ஓடு

● வினை ○ பெயர்

```
வினைச் சொற்கள்
**********************************************************************
ஓடு
ஓடினான்
ஓடினாள்
ஓடினார்
ஓடினார்கள்
ஓடினது
ஓடின
ஓடினாய்
ஓடினீர்
ஓடினீர்கள்
ஓடினேன்
ஓடினோம்
ஓடுகிறான்
ஓடுகிறாள்
ஓடுகிறார்
ஓடுகிறார்கள்
ஓடுகின்றது
ஓடுகின்றன
ஓடுகின்றாய்
```

Fig 3.1சொல் உருவாக்கி

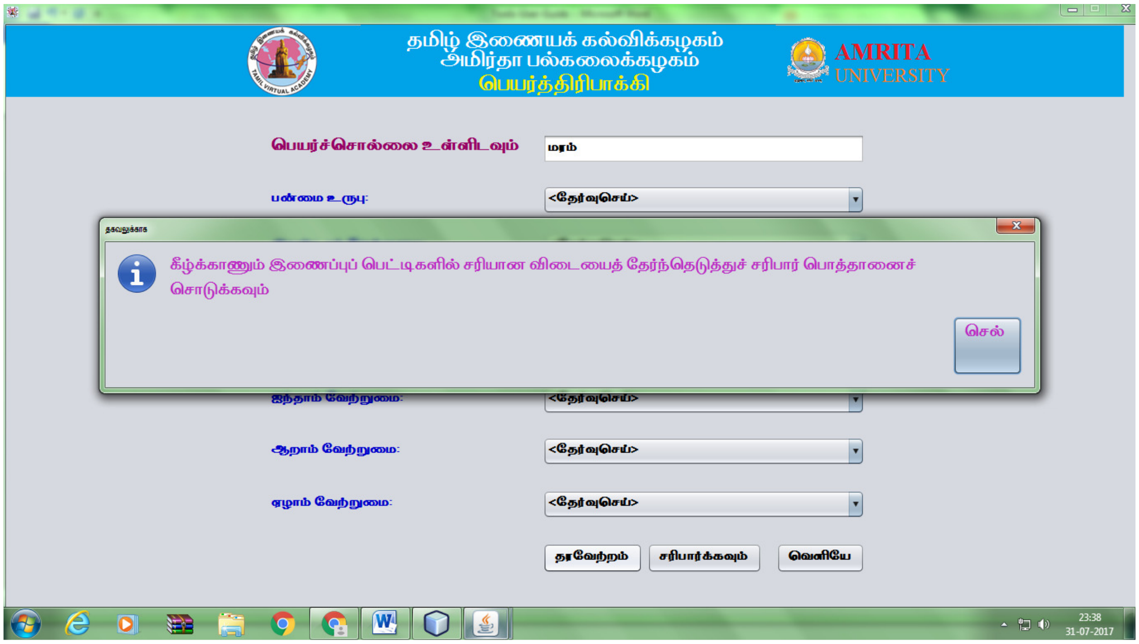
IV. Computing Tools for Teaching and Learning



பெயர்த்திரிபாக்கி

பெயர்ச்சொல்லை உள்ளிடவும்	
பன்மை உருபு:	<தேர்வுசெய்>
இரண்டாம் வேற்றுமை:	<தேர்வுசெய்>
மூன்றாம் வேற்றுமை:	<தேர்வுசெய்>
நான்காம் வேற்றுமை:	<தேர்வுசெய்>
ஐந்தாம் வேற்றுமை:	<தேர்வுசெய்>
ஆறாம் வேற்றுமை:	<தேர்வுசெய்>
ஏழாம் வேற்றுமை:	<தேர்வுசெய்>

Fig 4.1பெயர்த்திரிபாக்கி



தமிழ் இணையக் கல்விக்கழகம்
அமிர்தா பல்கலைக்கழகம்
பெயர்த்திரிபாக்கி

AMRITA UNIVERSITY

பெயர்ச்சொல்லை உள்ளிடவும்: மரம்

பன்மை உருபு: <தேர்வுசெய்>

தகவலுக்காக

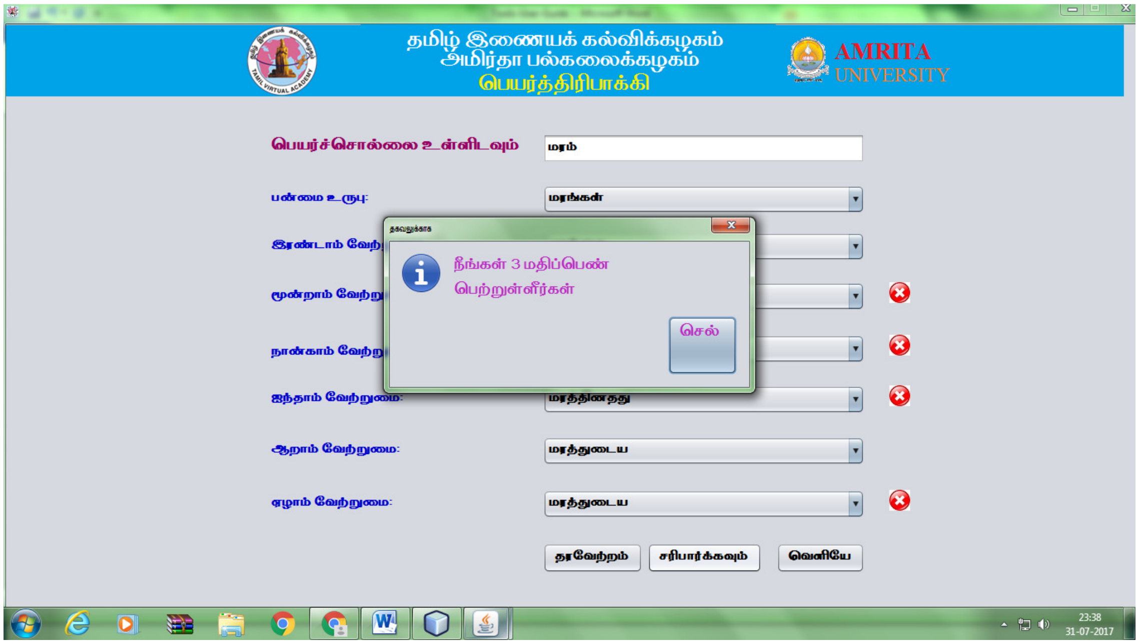
கீழ்க்காணும் இணைப்புப் பெட்டிகளில் சரியான விடையைத் தேர்ந்தெடுத்துச் சரிபார் பொத்தானைச் சொடுக்கவும்

ஐந்தாம் வேற்றுமை: <தேர்வுசெய்>

ஆறாம் வேற்றுமை: <தேர்வுசெய்>

ஏழாம் வேற்றுமை: <தேர்வுசெய்>

Fig 4.2பெயர்த்திரிபாக்கி



தமிழ் இணையக் கல்விக்கழகம்
அமிர்தா பல்கலைக்கழகம்
பெயர்த்திரிபாக்கி

AMRITA UNIVERSITY

பெயர்ச்சொல்லை உள்ளிடவும்: மரம்

பன்மை உருபு: மரங்கள்

இரண்டாம் வேற்

மூன்றாம் வேற்று ✗

நான்காம் வேற்று ✗

ஐந்தாம் வேற்றுமை: மரத்தினதது ✗

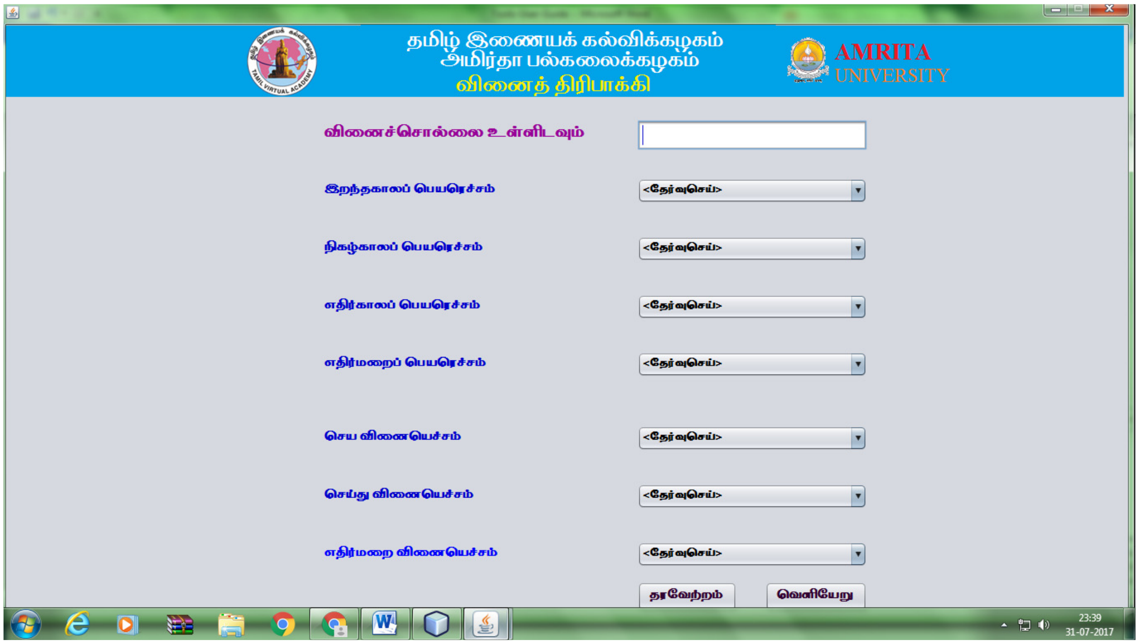
ஆறாம் வேற்றுமை: மரத்துடைய

ஏழாம் வேற்றுமை: மரத்துடைய ✗

தகவலுக்காக

நீங்கள் 3 மதிப்பெண் பெற்றுள்ளீர்கள்

Fig 4.3பெயர்த்திரிபாக்கி



தமிழ் இணையக் கல்விக்கழகம்
அமிர்தா பல்கலைக்கழகம்
வினைத் திரிபாக்கி

AMRITA UNIVERSITY

வினைச்சொல்லை உள்ளிடவும்

இறந்தகாலப் பெயரெச்சம் <தேர்வுசெய்>

நிகழ்காலப் பெயரெச்சம் <தேர்வுசெய்>

எதிர்காலப் பெயரெச்சம் <தேர்வுசெய்>

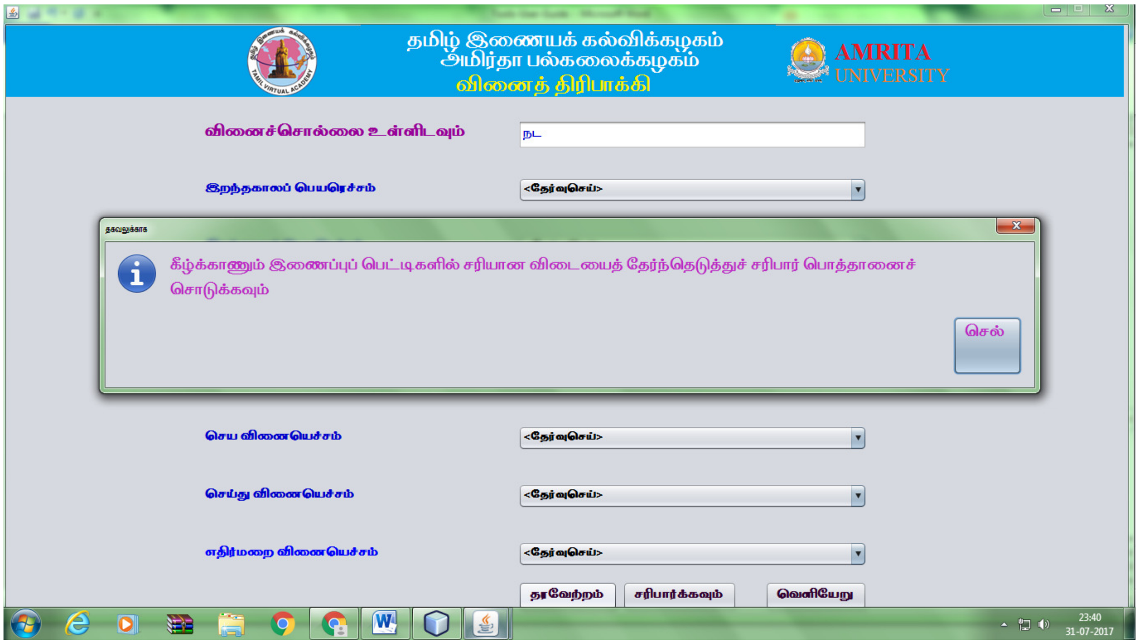
எதிர்மறைப் பெயரெச்சம் <தேர்வுசெய்>

செய வினையெச்சம் <தேர்வுசெய்>

செய்து வினையெச்சம் <தேர்வுசெய்>

எதிர்மறை வினையெச்சம் <தேர்வுசெய்>

Fig 4.4வினைத்திரிபாக்கி



தமிழ் இணையக் கல்விக்கழகம்
அமிர்தா பல்கலைக்கழகம்
வினைத் திரிபாக்கி

AMRITA UNIVERSITY

வினைச்சொல்லை உள்ளிடவும் நட

இறந்தகாலப் பெயரெச்சம் <தேர்வுசெய்>

தகவலுக்காக

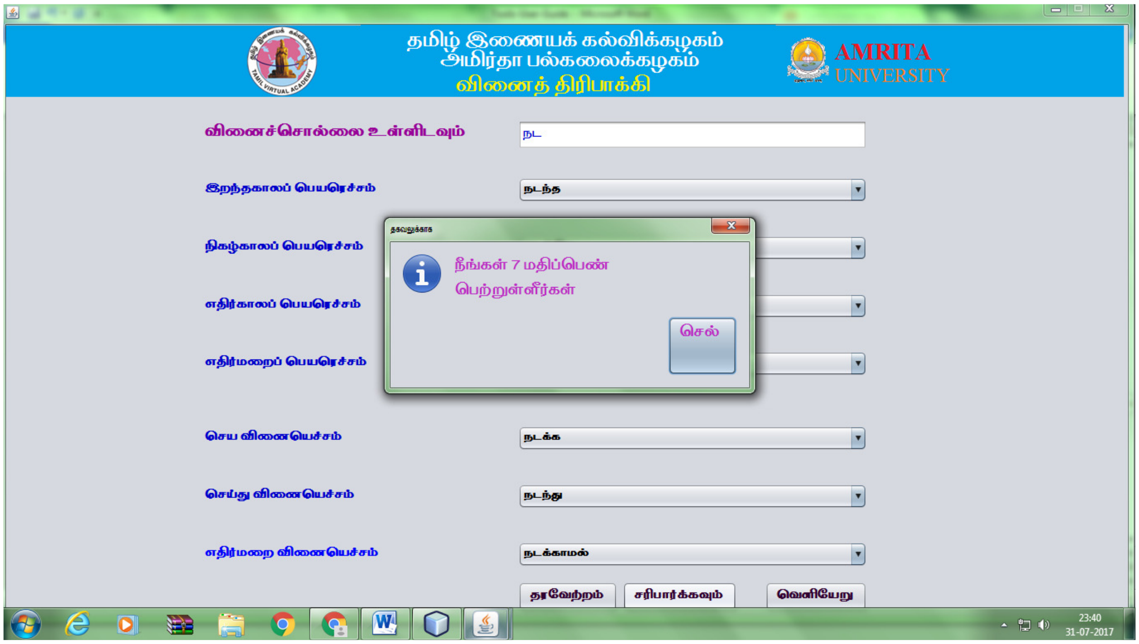
கீழ்க்காணும் இணைப்புப் பெட்டிகளில் சரியான விடையைத் தேர்ந்தெடுத்துச் சரிபார் பொத்தானைச் சொடுக்கவும்

செய வினையெச்சம் <தேர்வுசெய்>

செய்து வினையெச்சம் <தேர்வுசெய்>

எதிர்மறை வினையெச்சம் <தேர்வுசெய்>

Fig 4.5வினைத்திரிபாக்கி



தமிழ் இணையக் கல்விக்கழகம்
அமிர்தா பல்கலைக்கழகம்
வினைத் திரிபாக்கி

AMRITA UNIVERSITY

வினைச்சொல்லை உள்ளிடவும் நட

இறந்தகாலப் பெயரெச்சம் நடந்த

நிகழ்காலப் பெயரெச்சம்

எதிர்காலப் பெயரெச்சம்

எதிர்மறைப் பெயரெச்சம்

தகவலுக்காக

நீங்கள் 7 மதிப்பெண் பெற்றுள்ளீர்கள்

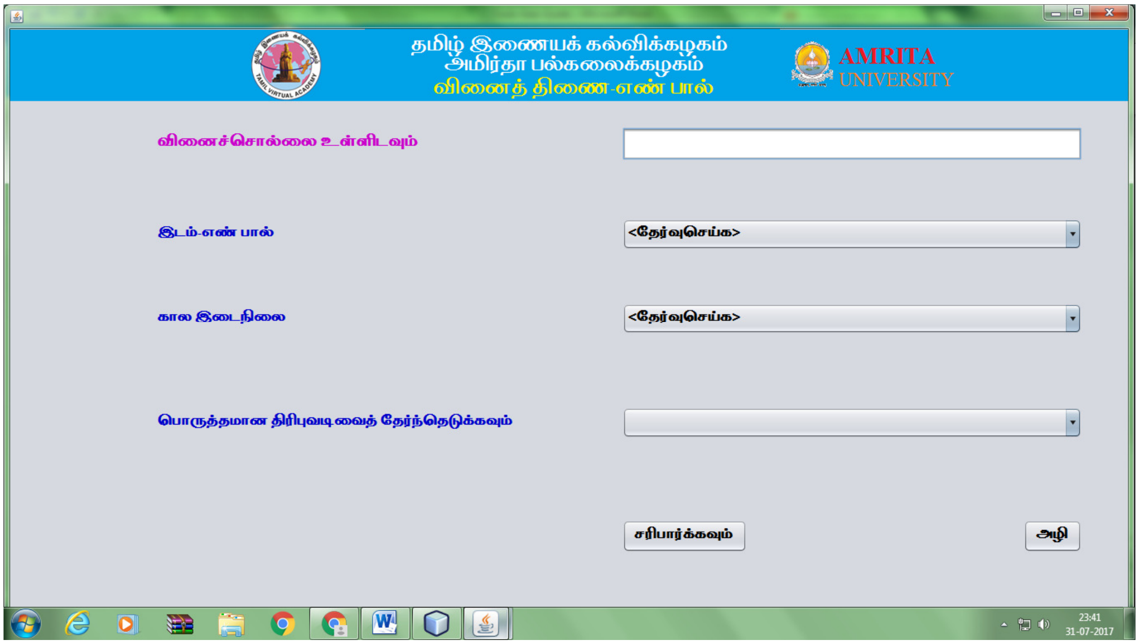
செல்

செய வினையெச்சம் நடக்க

செய்து வினையெச்சம் நடந்து

எதிர்மறை வினையெச்சம் நடக்காமல்

Fig 4.6வினைத்திரிபாக்கி



தமிழ் இணையக் கல்விக்கழகம்
அமிர்தா பல்கலைக்கழகம்
வினைத் திணை-எண் பால்

AMRITA UNIVERSITY

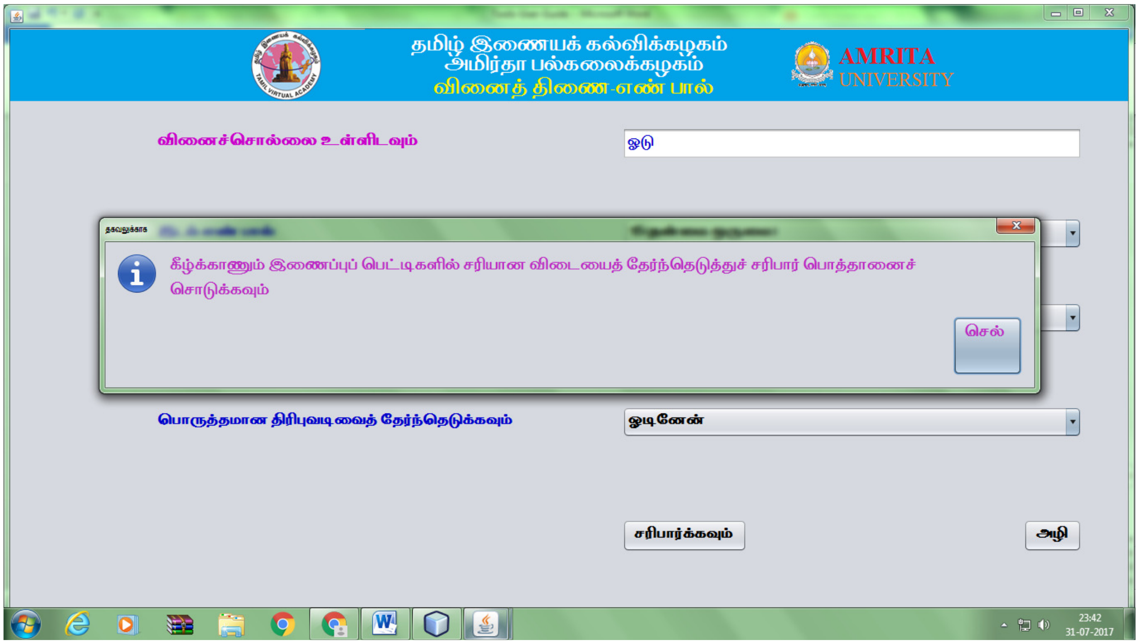
வினைச்சொல்லை உள்ளிடவும்

இடம்-எண் பால் <தேர்வுசெய்க>

கால இடைநிலை <தேர்வுசெய்க>

பொருத்தமான திரிபுவடிவைத் தேர்ந்தெடுக்கவும்

Fig 4.7வினைத் திை -எண் பால்



வினைத் திணை-எண் பால்

வினைச்சொல்லை உள்ளிடவும்: ஓடு

தகவலுக்காக

கீழ்க்காணும் இணைப்புப் பெட்டிகளில் சரியான விடையைத் தேர்ந்தெடுத்துச் சரிபார் பொத்தானைச் சொடுக்கவும்

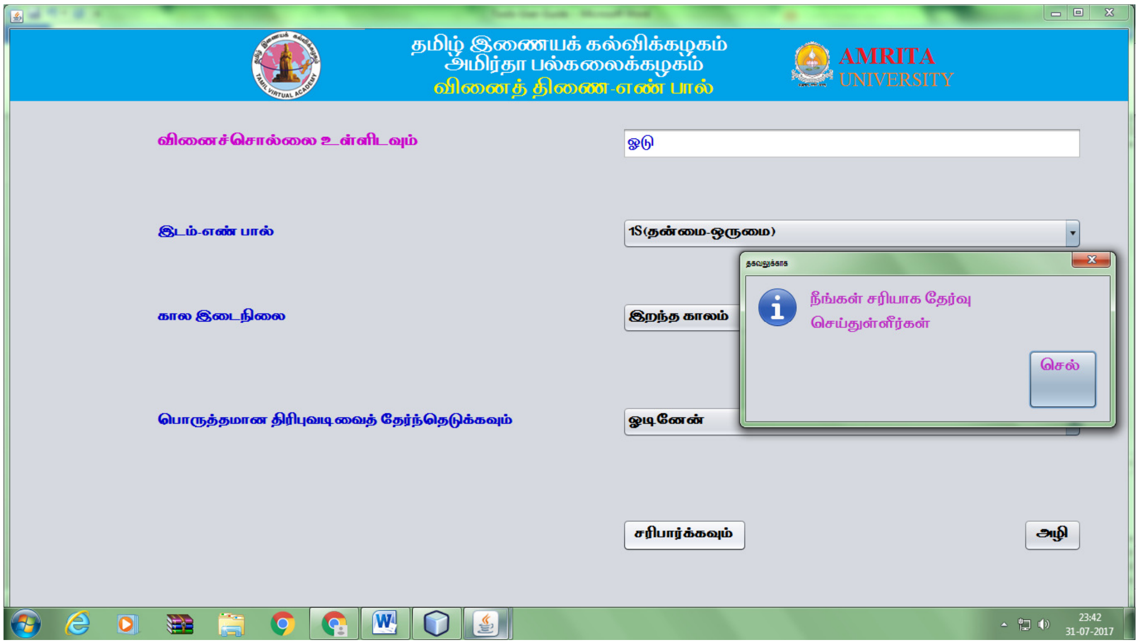
செல்

பொருத்தமான திரிபுவடிவைத் தேர்ந்தெடுக்கவும்: ஓடினேன்

சரிபார்க்கவும்

அழி

Fig 4.8வினைத் திை -எண் பால்



வினைத் திணை-எண் பால்

வினைச்சொல்லை உள்ளிடவும்: ஓடு

இடம்-எண் பால்: 1S(தன்மை-ஒருமை)

கால இடைநிலை: இறந்த காலம்

பொருத்தமான திரிபுவடிவைத் தேர்ந்தெடுக்கவும்: ஓடினேன்

தகவலுக்காக

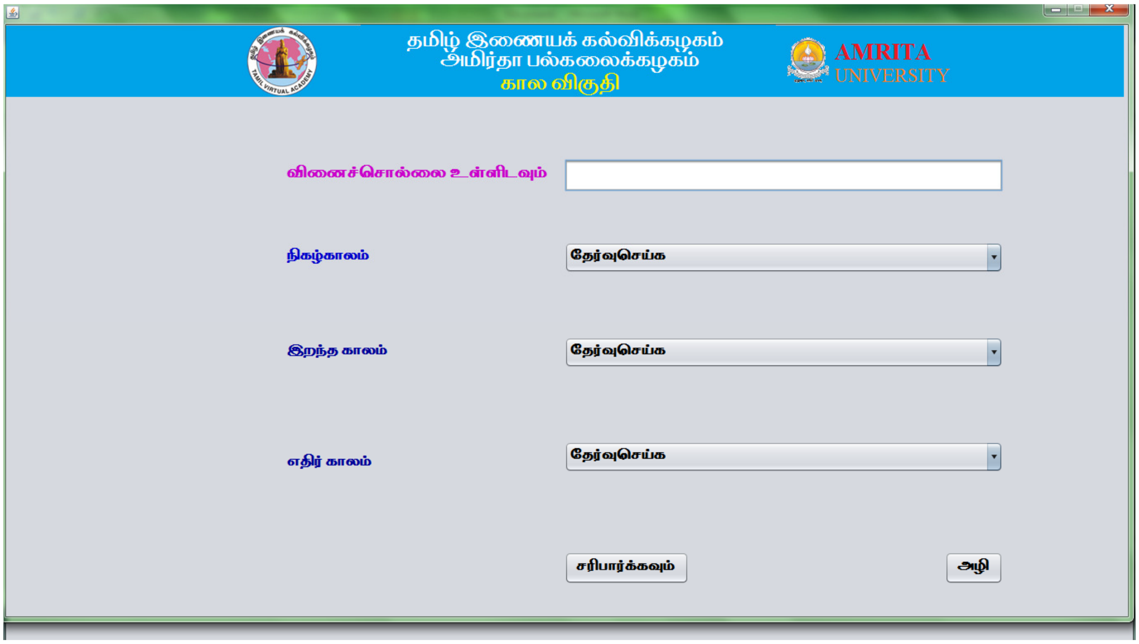
நீங்கள் சரியாக தேர்வு செய்துள்ளீர்கள்

செல்

சரிபார்க்கவும்

அழி

Fig 4.9வினைத் திை -எண் பால்



தமிழ் இணையக் கல்விக்கழகம்
அமிர்தா பல்கலைக்கழகம்
கால விகுதி

AMRITA
UNIVERSITY

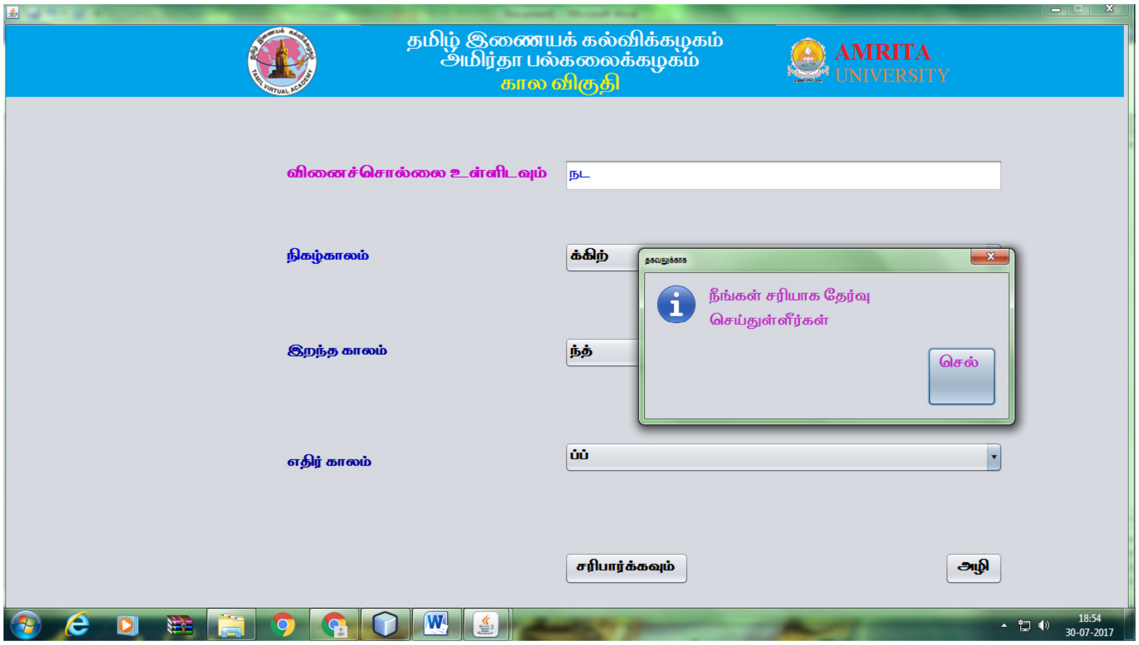
வினைச்சொல்லை உள்ளிடவும்

நிகழ்காலம்: தேர்வுசெய்க

இறந்த காலம்: தேர்வுசெய்க

எதிர் காலம்: தேர்வுசெய்க

Fig 4.10 கால விகுதி



தமிழ் இணையக் கல்விக்கழகம்
அமிர்தா பல்கலைக்கழகம்
கால விகுதி

AMRITA
UNIVERSITY

வினைச்சொல்லை உள்ளிடவும்: நட

நிகழ்காலம்: க்கிற்

இறந்த காலம்: ந்த்

எதிர் காலம்: ப்ப்

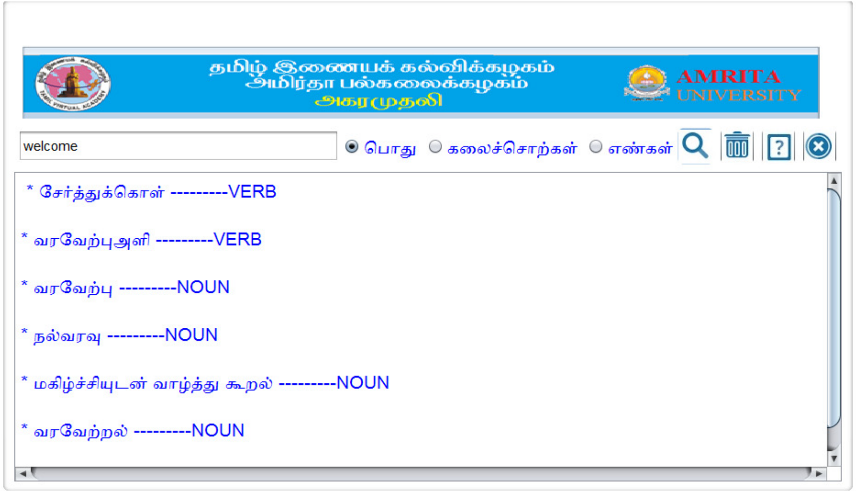
தகவலுக்காக

நீங்கள் சரியாக தேர்வு செய்துள்ளீர்கள்

Fig 4.11 கால விகுதி

A Smart English-Tamil Electronic dictionary

A Smart English-Tamil Electronic dictionary has been developed as the third component of the project. It has nearly one lakh entries for general vocabulary. The technical vocabulariesof various branches of knowledge have been integrated into the present super dictionary. The different domains of technical terms included in the super dictionary are the following: Administration, Agriculture, Arts and Humanities,Building and Construction, Economics, Engineering and Technology, Home Sciences, Law, Linguistics, Metallurgy, general Science, Social Work, Veterinary science, Chemical Engineering, Civil and Mechanical Engineering, Electrical and Electronics Engineering, Engineering and Technology(General), Information Technology, Mechanical Engineering, Textile Technology, Child Development, Cookery, First Aid Sick Nursing, Home Management, Home Science (General), Laundry, Needle Work , Botany, Chemistry, Geology, Mathematics, Physics, Statistics, and Zoology. A very useful GUI has been developed through which the user can get the information he expects from a English-Tamil bilingual dictionary. We hope to improve this tool with many user friendly and user useful components and develop it into a real smart English-Tamil dictionary.



welcome

● பொது ○ கலைச்சொற்கள் ○ எண்கள்

* சேர்த்துக்கொள் ---------VERB

* வரவேற்புஅளி ---------VERB

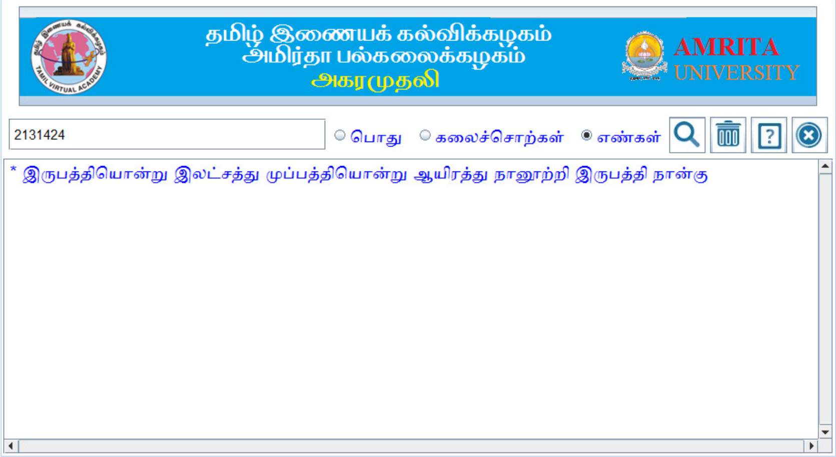
* வரவேற்பு ---------NOUN

* நல்வரவு ---------NOUN

* மகிழ்ச்சியுடன் வாழ்த்து கூறல் ---------NOUN

* வரவேற்றல் ---------NOUN

Fig 5.1அகரமுதலி



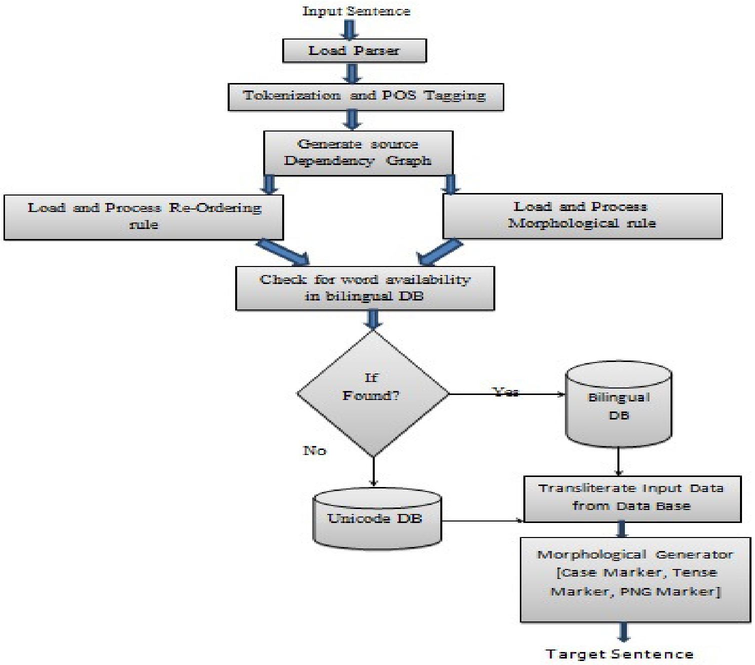
2131424

○ பொது ○ கலைச்சொற்கள் ● எண்கள்

* இருபத்தியொன்று இலட்சத்து முப்பத்தியொன்று ஆயிரத்து நானூற்றி இருபத்தி நான்கு

Fig 5.2 அகரமுதலி

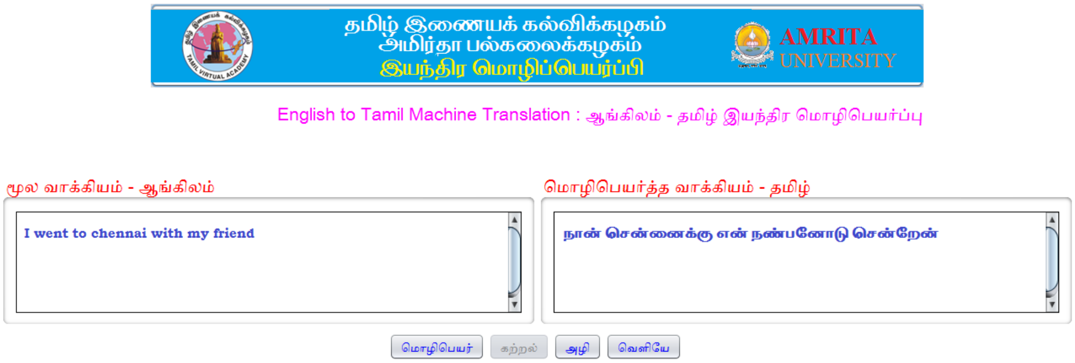
English to Tamil Rule Based Machine Translation System (RBMTS)


Input Sentence
Load Parser
Tokenization and POS Tagging
Generate source Dependency Graph
Load and Process Re-Ordering rule
Load and Process Morphological rule
Check for word availability in bilingual DB
If Found?
Yes
No
Bilingual DB
Unicode DB
Transliterate Input Data from Data Base
Morphological Generator [Case Marker, Tense Marker, PNG Marker]
Target Sentence

Flow Description:-

- The system accept input sentence from the user and passes to the Stanford parser to tokenize the sentence in to multiple word tokens and identify the POST(Parts of Speech Tagging) for each words from input sentence.
- After POST, Stanford parser generates dependency graph for source input sentence and then load both Re-Ordering rule and Morphological rule.
- After Morphological rule implementation, system check for each word token availability in Bilingual Database. If the word found, then transliteration process executed.
- If the word token not available in Bilingual Data Base, then system execute transliterate process from Unicode Data Base.
- After Transliteration, Morphological Generator generates and provides morphological information for input sentence. This process contains Case Marker, Tense Marker and PNG (Person, Number, and Gender) marker.
- Finally Morphological Generator generates Target Sentence.

Screenshot:-


தமிழ் இணையக் கல்விக்கழகம்
அமிர்தா பல்கலைக்கழகம்
இயந்திர மொழிப்பெயர்ப்பி
TAMIL VIRTUAL ACADEMY
AMRITA
UNIVERSITY
English to Tamil Machine Translation : ஆங்கிலம் - தமிழ் இயந்திர மொழிபெயர்ப்பு
மூல வாக்கியம் - ஆங்கிலம்
மொழிபெயர்த்த வாக்கியம் - தமிழ்
மொழிபெயர்
கற்றல்
அழி
வெளியே

I went to chennai with my friend

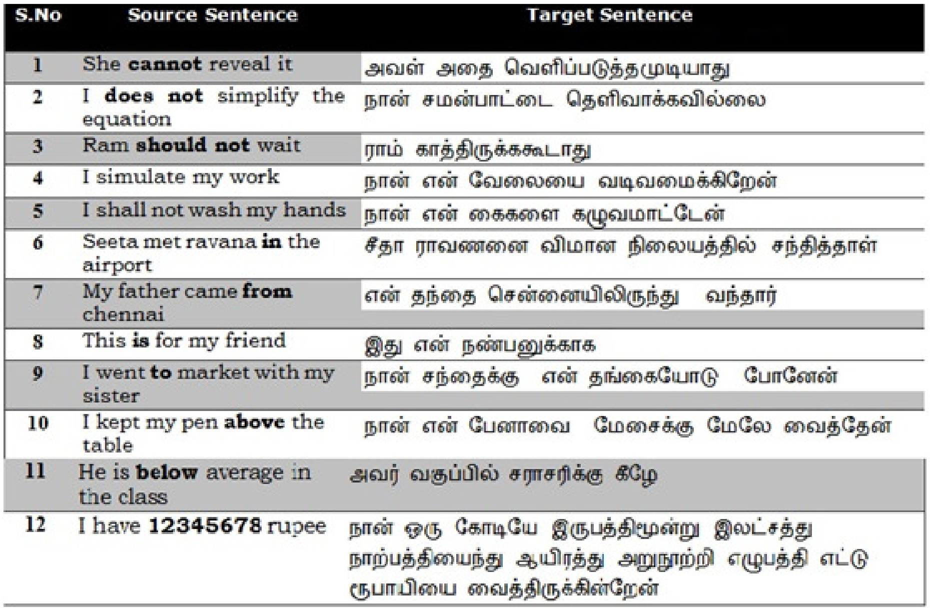
நான் சென்னைக்கு என் நண்பனோடு சென்றேன்

Fig 5.3 இயந்திர மொழிப்பெயர்ப்பி

Dictionary Statistics:-

✓	No. of Database	: 1
✓	No. of Table	: 1
✓	No.of records	: 9500
✓	Total Re-Ordering Rules	: 211
✓	Morphological Rules	: 29
✓	Parser Used	: stanford-parser3.5.2

Test Sentences:-



S.No	Source Sentence	Target Sentence
1	She **cannot** reveal it	அவள் அதை வெளிப்படுத்தமுடியாது
2	I **does not** simplify the equation	நான் சமன்பாட்டை தெளிவாக்கவில்லை
3	Ram **should not** wait	ராம் காத்திருக்ககூடாது
4	I simulate my work	நான் என் வேலையை வடிவமைக்கிறேன்
5	I shall not wash my hands	நான் என் கைகளை கழுவமாட்டேன்
6	Seeta met ravana **in** the airport	சீதா ராவணனை விமான நிலையத்தில் சந்தித்தாள்
7	My father came **from** chennai	என் தந்தை சென்னையிலிருந்து வந்தார்
8	This **is** for my friend	இது என் நண்பனுக்காக
9	I went **to** market with my sister	நான் சந்தைக்கு என் தங்கையோடு போனேன்
10	I kept my pen **above** the table	நான் என் பேனாவை மேசைக்கு மேலே வைத்தேன்
11	He is **below** average in the class	அவர் வகுப்பில் சராசரிக்கு கீழே
12	I have **12345678** rupee	நான் ஒரு கோடியே இருபத்திமூன்று இலட்சத்து நாற்பத்தியைந்து ஆயிரத்து அறுநூற்றி எழுபத்தி எட்டு ருபாயியை வைத்திருக்கின்றேன்

CONCLUSION AND FUTURE ENHANCEMENT

Each above stages are mandatory for translation engine, where even a single stages cannot be skipped to develop a simple system. Since the system approach us rule based, it performs well for all kind of simple sentences. But it may not able to translate exactly for long and complex sentences. So to overcome this difficulty, it should be extended and enhanced to the next level of machine translation methodology.