

Concept Space Synset Manager Tool

Apurva S Nagvenkar

DCST, Goa University

Taleigao Plateau, Goa.

apurv.nagvenkar@gmail.com

Neha R Prabhugaonkar

DCST, Goa University

Taleigao Plateau, Goa.

nehapgaonkar.1920@gmail.com

Venkatesh P Prabhu

Thyway Creation

Mapusa, Goa.

venkateshprabhu@thywayindia.com

Ramdas N Karmali

DCST, Goa University

Taleigao Plateau, Goa.

rnk@unigoa.ac.in

Jyoti D Pawar

DCST, Goa University

Taleigao Plateau, Goa.

jyotidpawar@gmail.com

Abstract

The IndoWordNet¹ Consortium consists of member institutions developing WordNet using the expansion approach.

The WordNets developed using expansion approach are very much influenced by the source language and may not reflect the richness of the target language (Walawalikar et al., 2010). And therefore the IndoWordNet Community decided to develop concepts which were specific to their respective language viz. language-specific concepts which will help in increasing the WordNet coverage. Besides the above requirement it was also felt that it should be possible to maintain additional information about the concepts i.e. an image, document describing the concept, links to websites and other resources, etc.

In this paper, we discuss a Concept Space Synset Management Tool (CSS)² which was developed to assist creation of language specific concepts/synsets and manage their linkages to other Indian language WordNets.

1 Background and Motivation

The IndoWordNet is a multilingual WordNet which links WordNets of different Indian languages on a common identification number

called as synset Id given to each concept (Bhattacharyya, 2010). WordNet is designed to capture the vocabulary of a language and can be considered as a dictionary cum thesaurus and much more (Miller, et al., 1993; Miller, 1995; Fellbaum, 1998).

Synset (Fellbaum, 1998) is composed of a gloss describing the concept, example sentences and a set of synonym words that are used for the concept. Besides synset data, WordNet maintains many lexical and semantic relations. Table1 gives the number of concepts/synsets created by the language groups of the Indradhanush WordNet Consortium which is a part of the IndoWordNet Consortium.

Sr. No.	Language	Nouns	Adjectives	Verbs	Adverbs	Total
1.	Bengali	27178	5183	3249	445	36685
2.	Gujarati	21659	5802	2804	444	30709
3.	Hindi	28163	6056	3079	456	37754
4.	Kashmiri	17959	6382	2354	305	27000
5.	Konkani	22912	5648	2983	471	32014
6.	Odia	27216	5273	2418	377	35284
7.	Punjabi	18982	5786	2808	442	28018
8.	Urdu	20816	5787	2800	443	29846

Table1: Synset linkage status

¹<http://www.cfilt.iitb.ac.in/indowordnet>

²<http://indradhanush.unigoa.ac.in/conceptspace>

Also a sense marked newspaper corpus (sense marking is a task to tag each word of the corpus with the WordNet sense) consisting of minimum 1,00,000 words has been created by each of the members of the Indradhanush WordNet Consortium. The coverage is found to be low. In order to increase the coverage of the WordNet it was decided that a corpus will be created by all language groups and the corpus will be sense marked.

To increase the coverage it was decided to add the concepts which were specific to their respective language viz. language-specific concepts and nullify the effect of influence of the source language on the target language WordNet. The CSS Manager Tool³ was developed to assist in creation of language-specific concepts, linking to other language WordNets, providing additional information about synsets, etc. The features and the detailed framework of the CSS Manager Tool is explained in section 3 and 4.

The rest of the paper is organized as follows – section 2 introduces the related work. The features of CSS Manager Tool are presented in section 3; section 4 presents the architecture of CSS Manager Tool. Section 5 presents the implementation details followed by the conclusion and future work.

2 Related Work

For many Indian languages, WordNets are constructed using the expansion model where Hindi WordNet synsets are taken as a source using the MultiDict Tool (Chatterjee, 2010) created by IIT Bombay. The tool also had feature to add comments and references but it was not an ideal tool for creation of language-specific synsets.

The limitations of the MultiDict Tool are:

- Creating and linking of language-specific synsets across languages was not possible,
- finding the overlap of synsets across languages was not possible,
- Feature to provide additional information about the synset was not present,
- Validation of synsets was not possible.
- Features to search synsets based on domain, date, category was not present.

And therefore the CSS Manager Tool was developed in order to overcome the above limitations.

3 Features of CSS Tool

CSS Manager Tool is a centralized tool meant for effective creation and management of synsets. The features supported currently by the CSS Manager Tool are as follows:

1. Synset Creation:

- Addition/update/validation of synsets, linking of two or more synsets with similar gloss across languages,
- Comments- Comments can be provided in case of any issue in the synset content.
- Allows adding additional information about the synset (images, documents, links, etc.).

2. Interactive User Interface:

- The CSS Manager Tool is designed keeping in mind the broadest range of users and contexts of use.
- Supports both left-to-right and right-to-left text rendition.
- Allows adjustment of the layout as per direction in which content language is written through a simple setting of a flag.
- Viewing various media added for clarity on synsets, etc.

3. Security:

- The CSS Manager Tool stores information in a centralized database system where access control mechanisms can more easily restrict access to your content.
- User Management supports adding/blocking/ unblocking users, and assigns privileges to the users.

4. Use of RBAC approach

- Role-based access control (RBAC) is an approach to restricting system access to authorized users.
- Roles are created for various functions. The permissions to perform certain operations are assigned to specific roles.
- Members or staff are assigned particular roles, and through those role assignments acquire the permissions to perform particular functions.
- Roles can be easily created, changed, or discontinued as the needs evolve, without having to individually update the privileges for every user.

³ <https://www.youtube.com/watch?v=BMhixBI7xOY&feature=youtu.be>

4 Architecture of CSS Tool

Figure 1 represents the architecture of CSS Manager Tool. The CSS Manager Tool is implemented in three blocks: User block, Super Admin block, and the Database. The CSS Manager tool is developed using the Hierarchical Role Based system with Access Control (RBAC) to control the access to certain parts and features of the CSS Manager Tool across different users. Refer Figure 2 for the block diagram of RBAC.

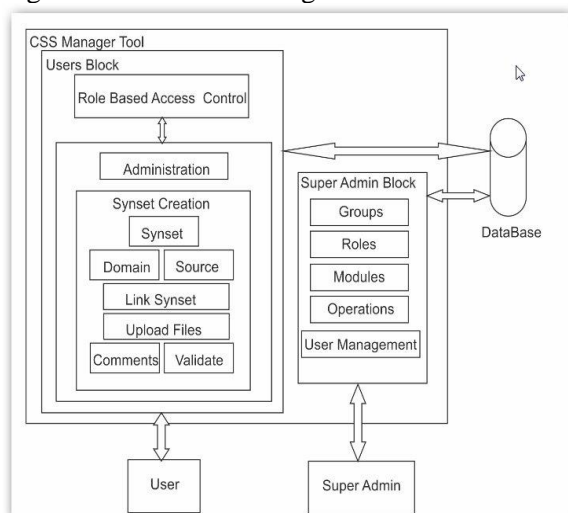


Figure1: Architecture of CSS Manager Tool

- The User block is responsible for creation/updation/validation of synsets, linking of synsets across languages, adding comments, source, and domain.
- The Super-Admin block is responsible for the creation of groups, users, roles to be assigned to the members in a group, modules and its operations, etc.
- The heart of the CSS Manager Tool is a centralized database that stores all the CSS data.

4.1 Modules of CSS Manager Tool

A module is an independent component which offers specific functionality. Each module is assigned different operations related to the module. The different operations are: Advance search, add/view/edit/delete/link synsets, and add/delete/change priority of example, add source, upload/delete file/add/view/reply comments, etc. Only those operations that need to be performed by members of a language group are assigned to the modules and these modules are allotted to the roles. These modules depend on CSS database. While the addition of new modules does not require any changes to the CSS database, new ta-

bles may need to be added to store data specific to module functionality.

Presently there are five modules, they are:

1. **View All Synset:** The view synset module allows the linguist to view synsets belonging to a language group/ category/ domain/source. The linguist/ lexicographer can perform the operations which are assigned for this module.
2. **Synset Creation:** Allows the linguist to create synsets. The linguist/ lexicographer can also add source/domain/images/ documents/links in order to give additional information about the synset.
3. **View Linked Synset:** Allows the linguist to view the list of synsets linked across languages.
4. **User Management:** Allows the administrator of a group to create new users, to block/unblock user, to assign privileges to the users, etc.
5. **Synset Validation:** Allows validation of synsets.

4.2 Role-Based system used in CSS Manager Tool

A role hierarchy is a way of organizing roles to reflect authority, responsibility, and competency. Some general operations may be performed by all the group members such as adding, viewing, searching synsets. In this situation, it would be inefficient and administratively cumbersome to specify repeatedly these general operations for each role that gets created. Therefore role hierarchy is used in order to avoid repetitive tasks. Also when a user is associated with a role, the user can be given additional privileges.

Currently, the CSS Manager Tool has four roles: Super admin, Admin, senior linguist and junior linguist.

- The super admin is responsible for creation of groups, users of a group, creation of roles to be assigned to the members in a group, addition of new modules and operations, and various other administrative operations such as adding source, domain, etc. which other roles cannot perform.

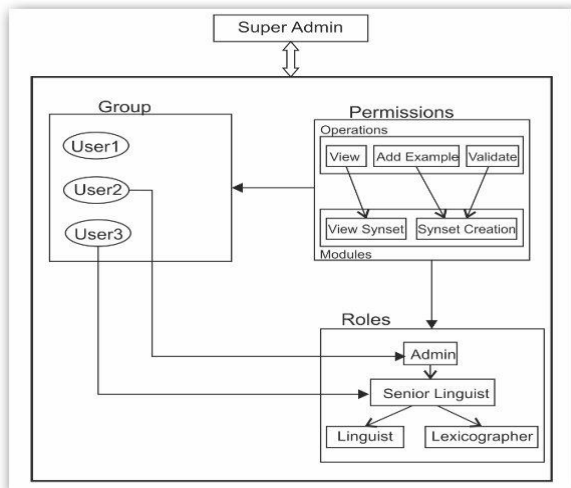


Figure2: Role Based system with Access Control

- The Admin is responsible for managing his/her language group created by the Super admin. The admin of a group can add/block users to his group. And can use all the modules which are assigned to the Admin by the Superadmin.
- The linguists are part of a language group. The operations (such as creating/validating/ linking of synsets) performed by the junior linguists are further validated and approved by the senior linguists of the group.

5 Implementation Details

The CSS Manager Tool is developed using PHP scripting language and is hosted on a Web Server supporting PHP version 5.3.15. Currently MySQL version 5.5.21 is used as database. The CSS Manager Tool was developed using XAMPP on 32 bit Microsoft Windows platform. It has been deployed on Fedora 16 Linux Platform using Apache version 2.2.22 and MySQL version 5.5.21 which come bundled with Fedora 16 Linux Platform. The screenshots of the tool are shown at the end of the paper.

6 Conclusion and Future Work

The advantages of CSS Manager Tool can be summarized as follows:

- **Ease in accessing synsets:** The synset is represented by an identification number called as synset id. Remembering id's is difficult for user, than remembering the concept of the synset. Earlier, the linguists had to remember synset id in order to perform any operation on synset in future. In CSS Manager Tool, the user

need not remember the synset ids, all the operations can be performed with the help of concept and synonymous set of the words.

- **Decentralized maintenance:** Need of specialized software or any specific kind of technological environment to access the tool is not required. Any browser device connected to the Internet would be sufficient for the job.
- **WordNet Enhancement:** Creation of language specific concepts/synsets, adding additional information about the synset and their linkages to other Indian language WordNets is possible. The tool is being enhanced to support validation of WordNets.

Acknowledgement

This work has been carried out as a part of the Indradhanush WordNet Project (11(13)/2010-HCC(TDIL), dated 3-8-2010) jointly carried out by nine institutions. We wish to express our gratitude to the funding agency DeitY, Govt. of India and also all the members of the Indradhanush Consortium.

References

- Pushpak Bhattacharyya. 2010. *IndoWordNet*, Lexical Resources Engineering Conference 2010 (LREC2010), Malta.
- Arindam Chatterjee, Salil Joshi, Mitesh Khapra, Pushpak Bhattacharyya. 2010. *Introduction to Tools for IndoWordNet and Word Sense Disambiguation*. 3rd IndoWordNet workshop, International Conference on Natural Language Processing.
- Christiane Fellbaum (ed). 1998. *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.
- George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine Miller. 1993. *Introduction to WordNet: An On-line Lexical Database*.
- George A. Miller. 1995. *WordNet: A Lexical Database for English*. Communications of the ACM Vol. 38, No. 11: 39-41.
- Shantaram Walawalikar, Shilpa Desai, Ramdas Karmali, Sushant Naik, Damodar Ghanekar, Chandrekha D'souza and Jyoti Pawar. 2010. *Experiences in Building the Konkani Word Net using the expansion Approach*. In Proceedings of the 5th GlobalWordNet Conference on Principles, Construction and Application of Multilingual WordNets (Mumbai-India).

Konkani WordNet: WordNet For Konkani Language:
<http://konkaniwordnet.unigoa.ac.in>

<http://indradhanush.unigoa.ac.in/conceptspace/>

IndoWordNet Website: Multilingual WordNet which links WordNets of eighteen Indian languages:
<http://www.cfilt.iitb.ac.in/indowordnet/>

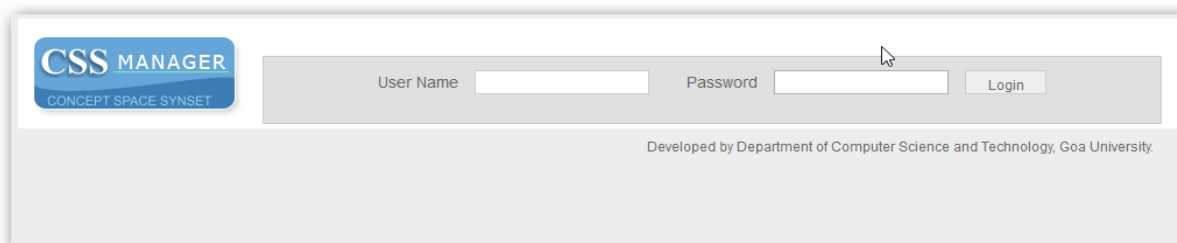
Concept Space Manager Tool Tutorial link:
<https://www.youtube.com/watch?v=BMhixBI7xOY&feature=youtu.be>

Indradhanush Website: WordNets for seven Indian Languages:
<http://indradhanush.unigoa.ac.in>

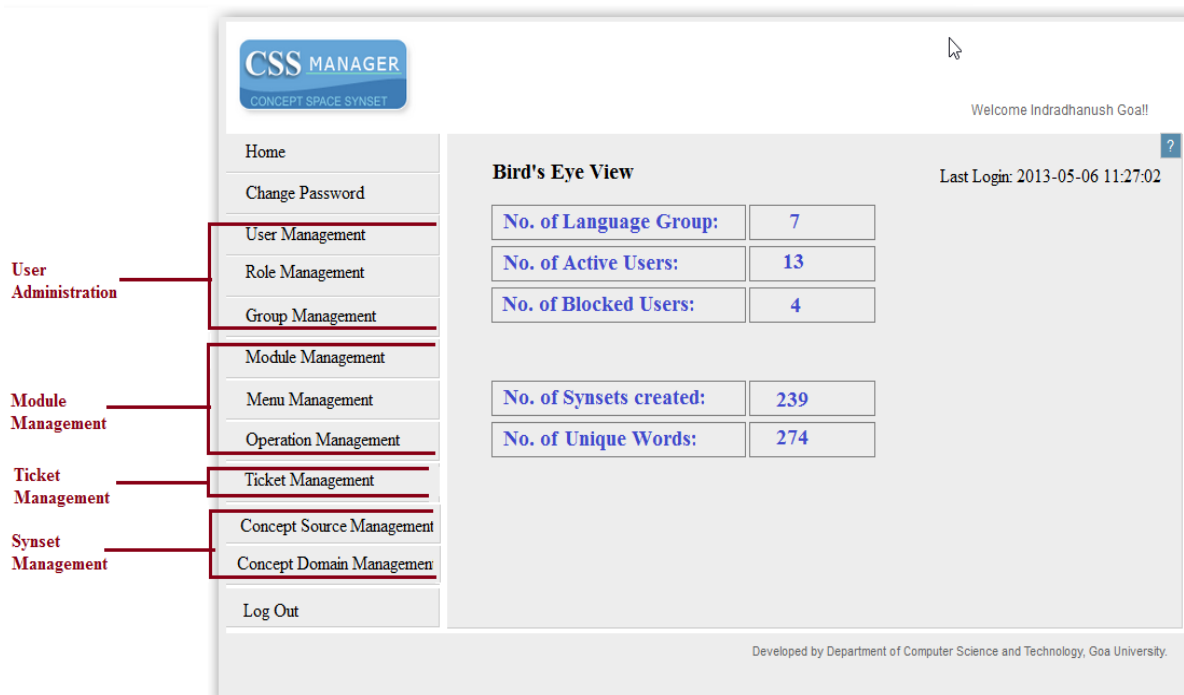
Concept Space Synset Manager Tool (CSS Manager Tool) :

Snapshots

1. **Login Page:** The login page of the CSS Manager Tool is shown below.



2. **SuperAdmin:** The super admin is the highest role in the role hierarchy. The super admin owns all the privileges which the admin, linguist or lexicographer have. The super admin is accountable for creation of groups, users of a group, creation of roles to be assigned to the members in a group, addition of new modules and operations, and various other administrative operations such as adding source, domain, etc. which other roles cannot perform. The snapshot of the super admin interface is shown below.



3. **User Management:** This module allows the administrator to view the users in a group, to add new users, to block or unblock user, to assign privileges to the users, etc. The User Management module is only available to the administrator of the group and not the linguist/ lexicographer.

CSS MANAGER
CONCEPT SPACE SYNSET

Welcome Indradhanush Goa

User Management

Add User

Search Users By :
Group : Konkani View By: Active Users Sort By: Latest Go

	User Name	email	Group	Contact Number
	Sanjay Singh	sanjay@gmail.com	Konkani	9881189299
	Sarvesh Nadkarni	sar@gmail.com	Konkani	9881189299
	Rashmi Shet	rashmi.1726@gmail.com	Konkani	7588921982
	Siddhi Kerkar	siddhi@gmail.com	Konkani	8976890431

Developed by Department of Computer Science and Technology, Goa University.

To add a new User,

CSS MANAGER
CONCEPT SPACE SYNSET

Welcome Indradhanush Goa

Add User

Name :

Address :

Birth Date :

Contact No :

Email :

Select Group : Konkani

Designation :

Is Active:

Password:

Confirm Password:

Save Cancel

Here, the Administrator of a group can add new user, block/unblock user, assign privileges to the user.

Developed by Department of Computer Science and Technology, Goa University.

The Modules which are available to the linguist and lexicographers are as follows:

- **Create Synset:** This module allows the user to create a new synset.

CSS MANAGER

CONCEPT SPACE SYNSET

Welcome Indradhanush Goa

Create Synset

View All Synsets

View Linked Synsets

Change Password

Log Out

Synset Creation

CSS ID 233

Concept Definition in Konkani

Concept Definition in Hindi

Example

Synsets

Domain

Source

Category Enable Keyboard

Transliterate Concept

Example in Konkani
 Example in Hindi

Add Konkani Example

Note: Synset should be comma(,) seperated


Upload File (image, link, pdf)

Upload File

Save

Save & Next

Developed by Department of Computer Science and Technology, Goa University.



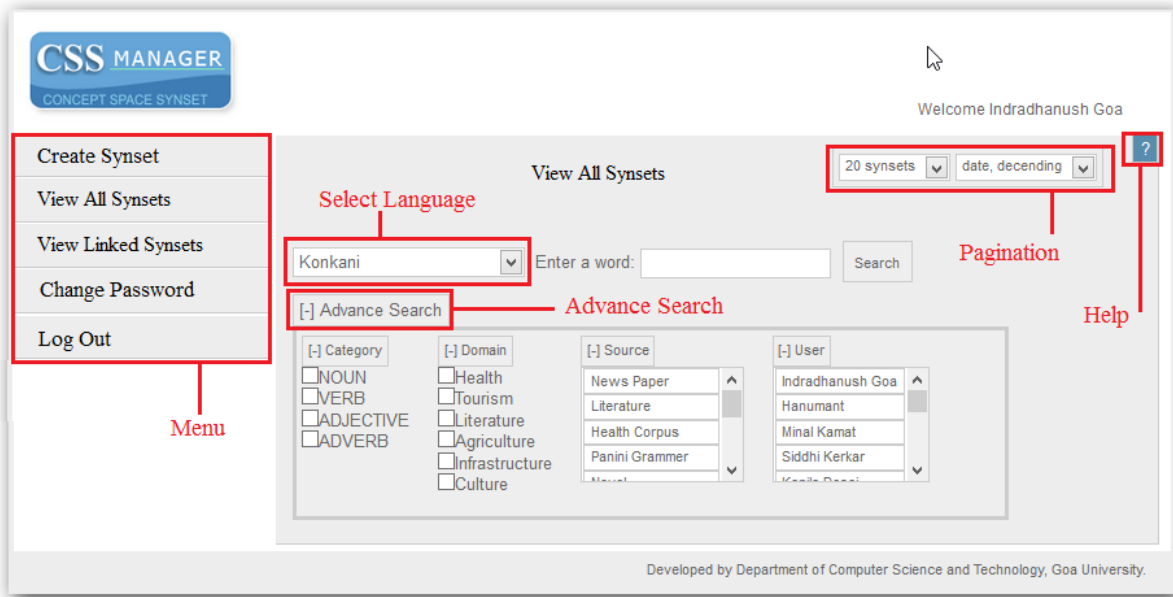
Lavani (Marathi: लावणी) is a genre of music popular in Maharashtra. According to a tradition, the word Lavani is derived from the word lavanya which means beauty. According to another tradition, it is derived from Marathi lavane.

People may not understand the meaning of the word "Lavani" based on its context. additional information about the concept in the form of image or doc or link can help the user to understand the concept.

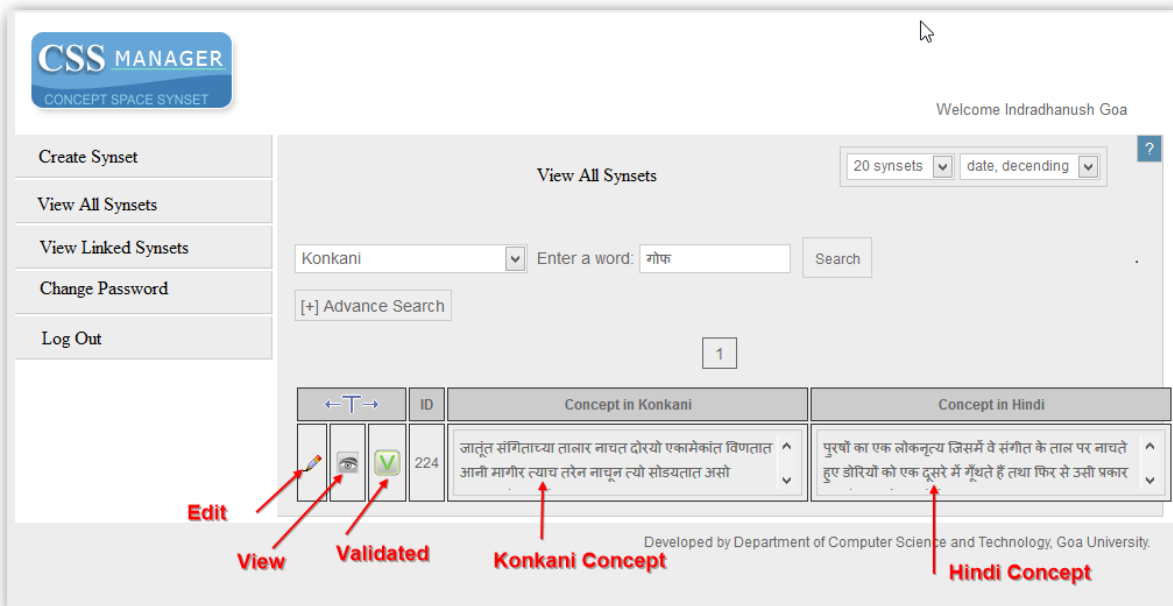
Lavani is performed by the female performers wearing nine-yard long saris. The songs are sung in a quick tempo.

- View All Synset:** This module allows the user to view all the synsets created so far. On selecting 'View All Synset' menu link, the user can view synsets belonging to a language. It also allows the user to select the number of synsets to be displayed per page, to view synsets based on the date of creation. Each module provides the user with the help files to assist in tool usage.

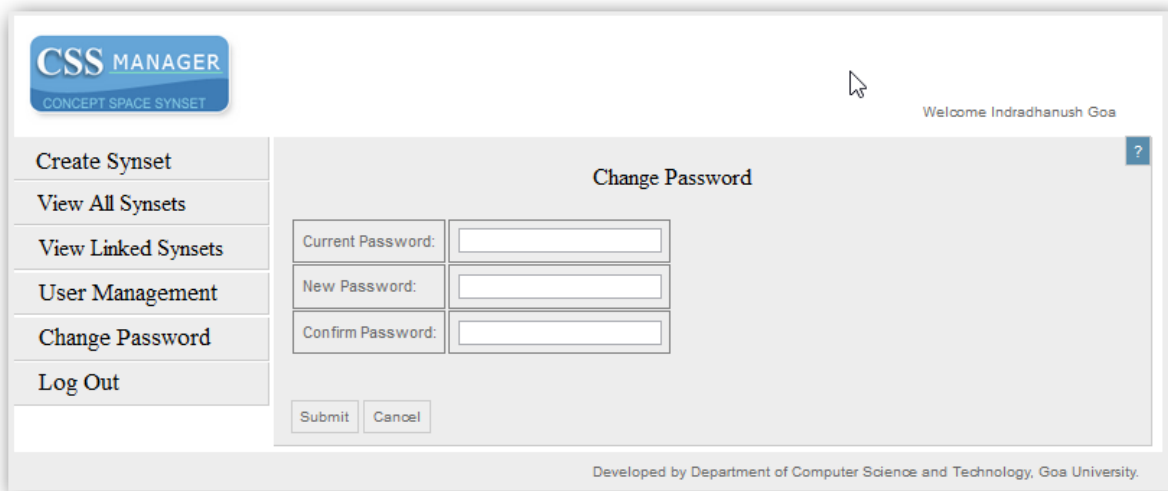
The 'Advance search' option allows the user to view synsets belonging to a particular grammatical category i.e Noun, Verb, Adverb, Adjective, a domain, a source and also to view the synsets created by a user of a group.



Based on the operations assigned to the modules and roles, the user can edit, view or validate the synsets.



- **View Linked Synsets:** This module is similar to the View All synset module, but it only allows the users to view the synsets which are linked across languages.
- **Change Password:** This module allows the user to change the password.



- **Log Out:** To log out from the CSS Manager Tool, the user needs to click on 'Log Out' from the menu list.