4.2 Digital Dictionary

4.2.1

◗ **Name of the Technology :** *Digital Dictionary (Bangla-Bangla)*

◗ **Nature of Technology :** Knowledge Resource (Digital Dictionary for Bangla to Bangla synonyms.)

◗ **Level:** (Product / Technology / Sub-system) Product

◗ **Technical Description of the Technology / Product including Basic block diagram, Algorithm used, O/S used, Front-end / user interface, and Specification of the Technology / Product :** At present, the Dictionary contains about 65,000 words with parts of speech and meaning. It shows instant meaning of word when spelling is known. Automatic searching and finding the meaning for unknown spelling is also possible. This software also supports meaning for euphonic and assimilated words (not available in the dictionary). Automatic search of words with particular character-combination/with particular parts of speech and other attributes is also possible. It generates a simplified Thesaurus using dictionary. It has the facility to add new words with meaning and parts of speech in the (personal) dictionary. This software is a useful tool to the researchers and common people.

Supporting platform(s): Windows 95/98/2000/NT

Memory requirements: 32MB or above

◗ **Representative Snapshot / screenshot of the Technology / Product:** Screenshot of Digital Dictionary:

◗ **Scalability / Portability / Expandability: Portability:** The system can run in WIN9X platform.

◗ **Readiness of Transfer of Technology (ToT) :** NA

◗ **Availability of documentation:** Yes

◗ **Testing of the Product / Technology:** Locally tested

◗ **IPR / Open-source:** Not open source

◗ **Potential Beneficiaries :** Typists, Linguists, Students, Teachers, Educationists, Office Personnel, Corpus Developers etc.

◗ **User–agency tie-up:** Solicited

◗ **Name and address of the Resource Person:**
B. B. Chaudhuri, Professor & Head,
Computer Vision & Pattern Recognition Unit
Indian Statistical Institute
203 Barrackpore Trunk Road, Calcutta 700 108,
INDIA
E-mail : bbc@isical.ac.in
Web : http://www.isical.ac.in/rc-bangla

4.2.2

◗ **Name of the Technology :** *ORIDIC (e-Dictionary (Oriya to English)*

◗ **Nature of the Technology:** Knowledge Resource

◗ **Level:** (Product / Technology / Sub-system): Product

◗ **Technical Description of the Technology / Product including Basic block diagram, Algorithm used, O/S used, Front-end / user interface, and Specification of the Technology / Product:** The e-Dictionary containing same information in electronic format has capacity to search any word within a moment. The fully user friendly X- Window has been designed, which facilitates different options up to user requirement. It provides the category, synonymy and corresponding English meaning of Oriya and English words. English transliteration allows users to type in phonetic English to get its corresponding Oriya character. Help option is also available to get the keystroke/keystrokes equivalent for each Oriya character in phonetic format. Search engine handles the misspelled words and gives some accurate suggestive words in both English and Oriya. The system is successfully running over 27000 Oriya and 20000 English words. This S/W is designed using Java and Java Swing (Front-end) for both the Windows-98/2000/NT and the Linux O/S.

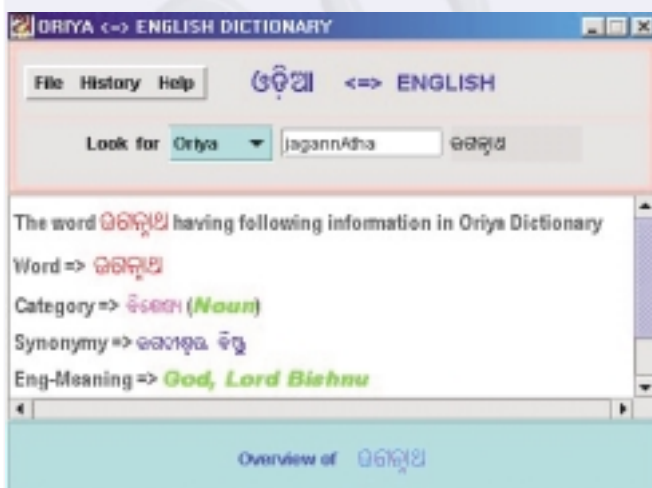◗ **Representative Snapshot / screenshot of the Technology / Product:** Product

Figure 1:- Output of the Oriya word SN_iÐ\ "jagannAtha" in e-Dictionary.

Figure 2:- Output of the English word "information" in e- Dictionary

◗ **Scalability / Portability / Expandability:** All

◗ **Readiness of Transfer of Technology (ToT):** Yes

◗ **Availability of documentation:** Yes

◗ Testing of the Product / Technology: Under Process

◗ **IPR / Open-source:** IPR (No.- SW-1175/2003)

◗ **Potential beneficiaries:** Common man

◗ **User-agency tie-up:** Solicited

◗ **Name and address of the Resource Person:** *Dr (Ms) Sanghamitra Mohanty RC-ILTS-Oriya, Dept. of CSA, Utkal University, Bhubaneswar, Orissa. E-mail : sanghami@rediffmail.com Web : http://www.ilts.utkal.org*

4.2.3

◗ **Name of the Technology :** *Trilingual (English-Hindi-Malayalam) Dictionary.*

◗ **Nature of Technology :** Knowledge Resource

◗ **Level :** Product

◗ **Technical Description :** An Online Trilingual (English-Hindi-Malayalam) Dictionary is developed. It Contains 50,000 plus words in each Language. English based Search and advanced search facilities are implemented. Search based on the other two languages being implemented.
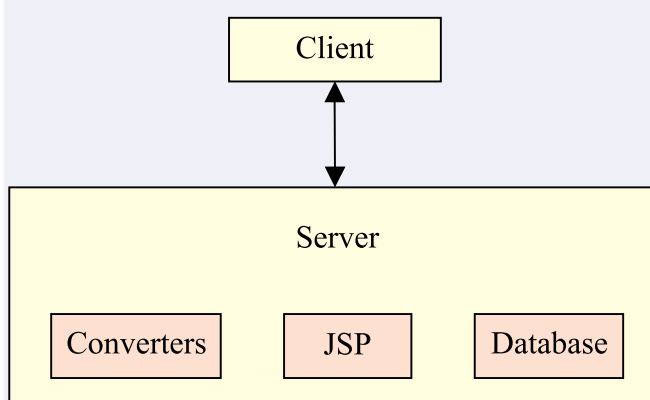

Figure 10.1 Block diagram of Trilingual Dictionary

The main features of the Dictionary are:

Portable in XML format.

ISCII Based.

Retrieval based on Parts Of Speech (POS)

Word search can be made in all three languages. (Implementation in progress)

Description of every word with example.

Advanced search facilities.

Extremely fast processing.

The Dictionary can be integrated with any other application or web portal. It can also be used as an aid for translation – both Machine aided and manual.

It helps the study of the concerned languages with relative ease.

◗ **Scalability/Portability/Expandability:** Switching to Unicode representation. Other languages can be added.

◗ **Readiness of Transfer of Technology (ToT):** Ready for TOT

◗ **Availability of documentation :** Documentation available.

◗ **Testing of the Product /Technology:** Development testing done.

◗ **Potential beneficiaries:** Teachers, Students, Linguist, etc.

◗ **User-agency tie-up:** Solicited

◗ **Name and address of the Resource Person:**
R. Ravindra Kumar
Additional Director & Coordinator
RCILTS-Malayalam,
*C-DAC, Thiruvananthapuram*
Email: ravi@erdcitvm.org.
Web : http://www.malayalamresourcentre.org

4.2.4

◗ Name of the Technology : *Sorputhaiyal (On-line Tamil Dictionary)*

◗ **Nature of Technology :** Content Development

◗ **Level :** Product

◗ **Technical Description :** Online Dictionary is a language-oriented software for the retrieval of lexical entries from monolingual and bilingual dictionaries by many users simultaneously. The Tamil dictionary contains 20,000 root words. It contains part of speech tag, Tamil and English synonyms for each word. The root words are classified into 15 categories.

Online Dictionary involves the following tasks.

1. Morphological analyser is used to get root word from inflected word before searching the dictionary

2. Searches the word based on syntactic category

3. Searches parts of word

4. Can also get Tamil equivalent words of given English words

Block Diagram

Block Diagram
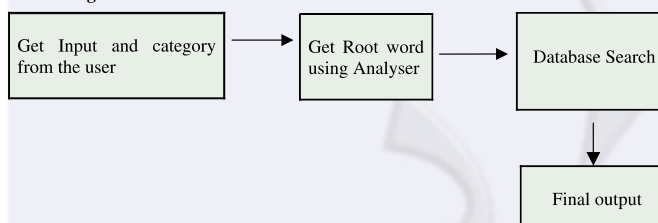
User Interface

The user interface provides a facility to the user to find Tamil words belonging to a category and all information for a given Tamil word. Each word with its type, Tamil and English meaning is displayed.

Specifications of the Technology:

The following environment is used for the development of Online Dictionary

Software Components

Online Dictionary was developed using Java, Servlets, Java Server Pages and MS-Access on Windows Operating System

Minimum Requirements to Run on Windows/ Linux

– Any Web server (Ex- Tomcat, JRun)

– Internet Explorer /Netscape Navigator

– J2RE 1.3

– Ram 128 MB

– Any Tamil Keyboard driver (Font encoding in TAB)

– Code converter for other font encounters is provided.

◗ **Representative Snapshots:**

◗ **Scalability:** Yes

◗ **Portability:** Yes

◗ **Readiness of Transfer of Technology :** Yes

◗ **Availability of Documentation:** User and Technical Manuals are available

◗ **Testing of the Product :** Tested for many words in the dictionary. Tested by many users during conference.

◗ **Open Source :** Yes

Potential Beneficiaries

Public

◗ **User Agency Tie up :** Solicited

◗ **Name and address of the Resource Person:** *Dr. T.V.Geetha, Dr.Ranjani Parthasarathy, RCILTS Tamil, Anna University, Chennai-600025 E-mail : rp@annauniv.edu Web : http://www.rs.anna.univ.edu*

4.2.5

◗ **Name of the Technology :** *English to Punjabi Official language Dictionary*

◗ **Nature of Technology :** Knowledge Resource

◗ **Level:** (Product / Technology / Sub-system) Product

◗ **Technical Description of the Technology / Product including Basic block diagram, Algorithm used, O/S used, Front-end / user interface, and Specification of the Technology / Product:** A technical glossary of around 17,000 English – Punjabi administrative terms have been developed and have also been uploaded on the internet. A CD has also been developed and the glossary has been installed on computers in many offices of the Punjab State Government.

– Platform: IBM Compatible PC with 486 or

– Pentium CPU

– Operating System: Windows 95,98,2000, XP

– Programming Interface: Delphi

◗ **Representative Snapshot / screenshot of the Technology / Product:**
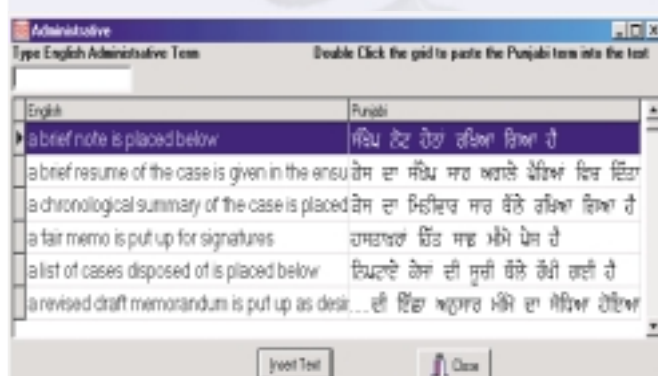
Screenshots of Online Punjabi Shabdkosh – Punjabi English Administrative Terms
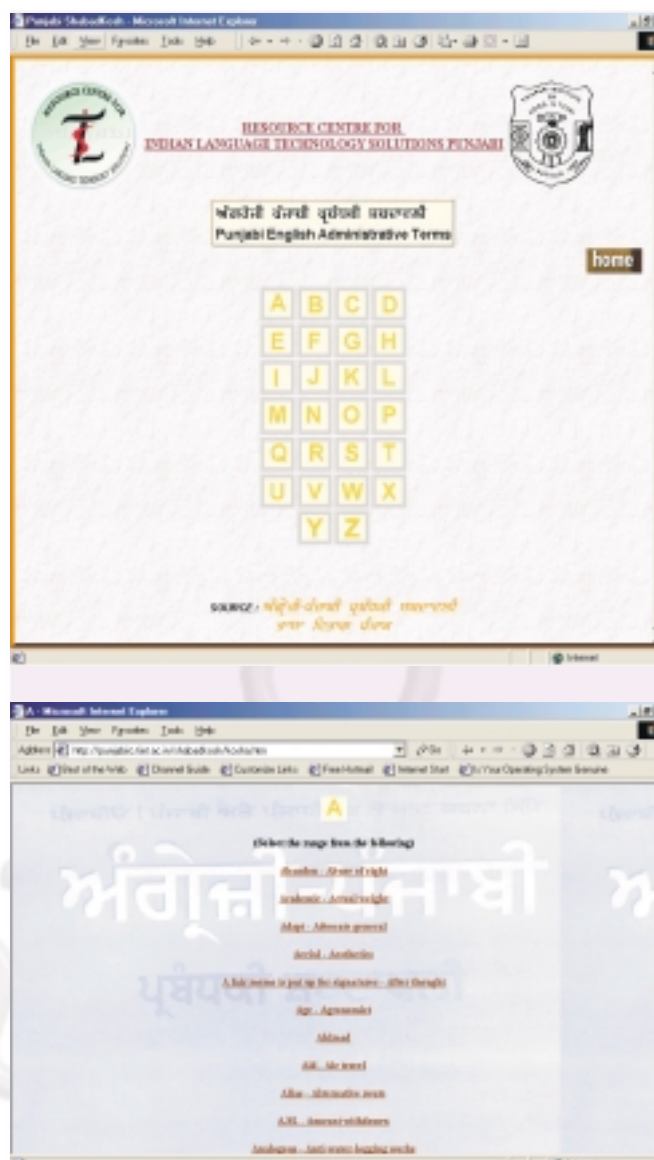
◗ **Scalability / Portability / Expandability :** Portable among any of the Windows environment Win 95, 98, Window-ME, Win-2000, Win-NT.

◗ **Readiness of Transfer of Technology (ToT):**

◗ **Availability of documentation :** Yes

◗ **Testing of the Product / Technology:** Testing done

◗ **IPR / Open-source:** Available in open source as the online version of the dictionary on the Resource Centre website at: http://punjabirc.tiet.ac.in/shabadkosh/shabadkosh.html

◗ **Potential Beneficiaries :** Typist, Linguists, Students, Teachers, Punjabi Learners, Translators, Educationists, Office Personnel, Corpus Developers etc.

◗ **User–agency tie-up** : Solicited

◗ **Name and address of the Resource Person:**
Dr. R.K. Sharma, Coordinator,
Resource Centre for Indian Languages Technology
Solutions – Punjabi (RCILTS - Punjabi),
School of Mathematics and Computer Applications
(SMCA),
T.I.E.T, Patiala - 147004 Punjab (India).
E-mail : rksharma@tiet.ac.in
Web : http://www.punjabirc.tiet.ac.in/rcentrei.asp

4.2.6
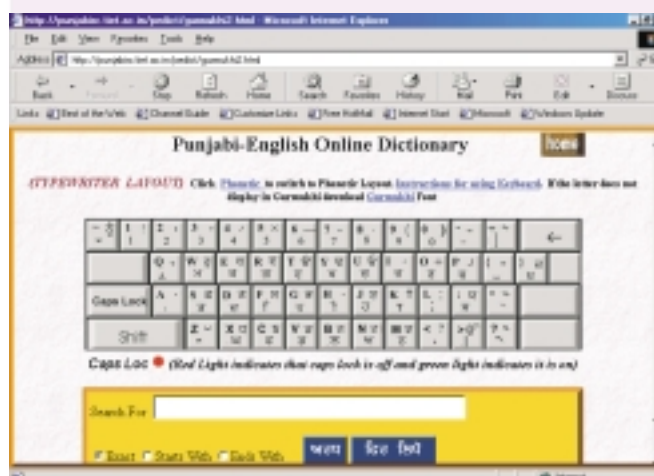
◗ **Name of the Technology :** *Online Dictionary (Punjabi to English)*

◗ **Nature of Technology :** Knowledge Resource (Software for Punjabi Language)

◗ **Level:** (Product / Technology / Sub-system) Technology

◗ **Technical Description of the Technology / Product including Basic block diagram, Algorithm used, O/S used, Front-end / user interface, and Specification of the Technology / Product:** The Online Dictionary has about 39,000 Punjabi words. Also the dictionary has sample sentences for common words and clips of pronunciation The dictionary is accessible in both typewriter and phonetic layout. One can search for a complete string match or a pattern in the dictionary. Sorting algorithm for Gurmukhi words has been used to sort the Gurmukhi words in the database. Also the feature of Dynamic fonts has been used in the dictionary.

Web Server: IIS 5.0

Operating System: Windows 2000 Advanced Server.

Programming Interface: ASP, JavaScript, HTML

◗ **Representative Snapshot / screenshot of the Technology / Product:**

◗ **Scalability / Portability / Expandability :** The dictionary is best viewed with Internet Explorer4.0, 5.0 with screen resolution 800/600.

◗ **Readiness of Transfer of Technology (ToT):** Yes

◗ **Availability of documentation :** Yes

◗ **Testing of the Product / Technology :** Testing of the dictionary will be over by September 2003.

◗ **IPR / Open-source:** Available as Open source on the Resource Centre website at:

http : //punjabirc.tiet.ac.in/pedict/gurmukhi2.html

◗ **Potential Beneficiaries :** Punjabi learners, Punjabi Typists, Language Experts, Writers.

◗ **User–agency tie-up :** Solicited

◗ **Name and address of the Resource Person:**
Dr. R.K. Sharma, Coordinator,
Resource Centre for Indian Languages Technology
Solutions – Punjabi (RCILTS - Punjabi),
School of Mathematics and Computer Applications
(SMCA),
T.I.E.T, Patiala - 147004 Punjab (India).
E-mail : rksharma@tiet.ac.in
Web : http://www.punjabirc.tiet.ac.in/rcentrei.asp

4.2.7

◗ **Name of the Technology : Online Dictionary (Hindi to Punjabi)**

◗ **Nature of Technology :** Knowledge Resource (Software for Punjabi Language)

◗ **Level :** (Product / Technology / Sub-system) Technology

◗ **Technical Description of the Technology / Product including Basic block diagram, Algorithm used, O/S used, Front-end / user interface, and Specification of the Technology / Product :** The Online Dictionary has about 45,000 words. The dictionary displays the words in alphabetical order. Also the feature of Dynamic fonts has been used in the dictionary.

Web Server: IIS 5.0

Operating System: Windows 2000 Advanced Server.

Programming Interface: HTML/JavaScript

◗ **Representative Snapshot / screenshot of the Technology / Product:**

◗ **Scalability / Portability / Expandability :** The dictionary is best viewed with Internet Explorer 4.0, 5.0 with screen resolution 800 x 600.

◗ **Readiness of Transfer of Technology (ToT):**

◗ **Availability of documentation :** Yes

◗ **Testing of the Product / Technology:** Testing of the dictionary will be over by September 2003.

◗ **IPR / Open-source :** Available as Open source on the Resource Centre website at:

http : //punjabirc.tiet.ac.in/hindidict/hindidictinput.htm

◗ **Potential Beneficiaries :** Punjabi learners, Punjabi Typists, Language Experts, Writers.

◗ **User–agency tie-up :** Solicited

◗ **Name and address of the Resource Person :**
Dr. R.K. Sharma, Coordinator,
Resource Centre for Indian Languages Technology
Solutions – Punjabi (RCILTS - Punjabi),
School of Mathematics and Computer Applications
(SMCA),
T.I.E.T, Patiala - 147004Punjab (India).
E-mail : rksharma@tiet.ac.in

4.2.8

◗ **Name of the Technology :** *Online Dictionary (English to Punjabi)*

◗ **Nature of Technology :** Knowledge Resource (Software for Punjabi Language)

◗ **Level:** (Product / Technology / Sub-system) Technology

◗ **Technical Description of the Technology / Product including Basic block diagram, Algorithm used, O/S used, Front-end / user interface, and Specification of the Technology / Product :** The Online Dictionary has about 46,000 words. One can search for a complete string match or a pattern in the dictionary. Also the feature of Dynamic fonts has been used in the dictionary.

Web Server: IIS 5.0

Operating System: Windows 2000 Advanced Server.

Programming Interface: ASP, JavaScript, HTML

◗ **Representative Snapshot / screenshot of the Technology / Product:**

◗ **Scalability / Portability / Expandability :** The dictionary is best viewed with Internet Explorer 4.0, 5.0 with screen resolution 800 x 600.

◗ **Readiness of Transfer of Technology (ToT):**

◗ **Availability of documentation :** Yes

◗ **Testing of the Product / Technology:** Testing of the dictionary will be over by September 2003.

◗ **IPR / Open-source :** Open Source

◗ **Potential Beneficiaries :** Punjabi learners, Punjabi Typists, Language Experts, Writers.

◗ **User–agency tie-up:**

◗ **Name and address of the Resource Person :**
*Dr. R.K. Sharma, Coordinator,
Resource Centre for Indian Languages Technology Solutions – Punjabi (RCILTS - Punjabi),
School of Mathematics and Computer Applications (SMCA),
T.I.E.T, Patiala - 147004 Punjab (India).
E-mail : rksharma@tiet.ac.in
Web : http://www.punjabirc.tiet.ac.in/rcentrei.asp*

4.2.9

◗ **Name of the Technology :** *Telugu Dictionaries*

(i) English - Telugu Dictionary, (ii) Telugu - English Dictionary, (iii) Telugu - Hindi Dictionary

◗ **Nature of the Technology :** Knowledge Resource

◗ **Level :** Resource required for machine translation system

◗ **Technical Description**

(i) C P Brown's English - Telugu Dictionary
Status : Completed; Size : 31,000 plus; Fields : POS, Meanings, Usage; XML : Yes; Indexed : Yes; Webenabled : Yes.

(ii) C P Brown's Telugu - English Dictionary
Status : Completed; Size : 31,000 plus;
Fields : POS,Meanings, Usage, Etymology
XML : Work In progrss Web-enabled? : Work in progress.

(iii) Telugu - Hindi Dictionary
Status : Completed; Size : 64,000 plus;
Fields : POS, Paradigm Class, Meanings
XML : Yes; Indexed : Yes; Web- enabled : Work in progress

◗ **Operating System used :** LINUX

◗ **Scalability / Portability / Expandability :** Available

◗ **Availability of documentation:** Yes

◗ **Testing of the Product / Technology :** Testing has been done

◗ **IPR / Open Source :** IPR lies with University of Hyderabad and Dept of Information Technology

◗ **Potential Beneficiaries :** Telugu Software developers, Liguists, teachers, students etc.

◗ **User agency tie-up :** Solicited

◗ **Name and address of the Resource Person :**
Dr. K. Narayanamurthy
Department of CIS, University of Hyderabad
Hyderabad-500046, AP, India
E-mail : knmcs@uohyd.ernet.in
Web : http://www.languagetechnologies.ac.in

4.2.10

◗ **Name of the Technology :** Hindi Speech Data base.

◗ **Nature of Technology :** Knowledge Resources-Speech Database

◗ **Level: (Product / Technology / Sub-system) :** Sub-system

◗ **Technical Description of the Technology / Product including Basic block diagram, Algorithm used, O/S used, Front-end / user interface, and Specification of the Technology / Product :** As part of VOICE-II project, a labeled and time-aligned database of spoken Hindi sentences has been generated with the collaborative efforts of the speech groups at TIFR, Mumbai and CEERI, New Delhi. Although the intended principal end-use of this multi-speaker, continuous speech database is in training of a speech recognition system for Hindi, the design specifications of the database are so general that it aids tasks such as speaker recognition, study of acoustic-phonetic correlates of the language. speaker variabilities in Hindi speech. For further details, the interested readers may refer to the following publications:

Samudravijaya K, K.D.Rawat, and P.V.S.Rao, "Design of Phonetically Rich Sentences for Hindi Speech Database", J. Ac. Soc. Ind. vol. XXVI, December 1998, pp. 466-471.

◗ **Representative Snapshot / screenshot of the Technology / Product:** Not Available

◗ **Scalability / Portability / Expandability:** This can be used in any Window or Unix system.

◗ **Readiness of Transfer of Technology (ToT):** The CD of the data base are available. This was handed over to the Department of Information Technology.

◗ **Availability of documentation:** Brief description and user guide is available.

◗ **Testing of the Product / Technology:** The product can be tested only in conjunction with the speech recognition systems.

◗ **IPR / Open-source:** As the CD has been handed over to the MICT, it depends on them.

◗ **Potential beneficiaries:** Researchers of speech technology

◗ **User-agency tie-up:** Nil

◗ **Name and address of the Resource Person:**
Dr. K. Samudravijaya,
School of Technology and computer Science,
Tata Institute of Fundamental Research,
Homi Bhabha Road, Mumbai 400 005