Department of Information Technology

PROFORMA FOR SUBMITTING R&D PROJECT PROPOSAL
FOR SEEKING FINANCIAL SUPPORT

27th January 2011

SUMMARY SHEET

1. **Title of Project:** Socionity: Web with Community Sourcing of Local Content

2. **Organization:**

   a) Name: International Institute of Information Technology- Hyderabad
   b) Address: Gachibowli, Hyderabad- 500 032
      Andhra Pradesh, INDIA
      Phone : +91-40-6653 1000
   c) Legal status (indicate if Government Department, Statutory, Corporate Body, Registered Society, Private Company with recognised R&D unit etc.): Deemed University

3. **3.1 Chief Investigator:**

   a) Name: Prof. Navjyoti Singh
   b) Designation: Professor
   c) Department: Centre for Exact Humanities
      Address: International Institute of Information Technology
      Gachibowli, Hyderabad 500 032
      Andhra Pradesh, INDIA
      Phone : +91-40-6653 1403, 6653 1302

   3.2 Co-Investigator:

   a) Name: Dr. Rajeev Sangal
   b) Designation: Professor & Director
   c) Departmnent: Language Technologies Research Centre
      Address:International Institute of Information Technology
      Gachibowli, Hyderabad 500 032
      Andhra Pradesh, INDIA
      Phone:+91-40-23001412

4. **Nature of Project (Check one)**

   a) **Research, Development & Engineering (R,D & E) leading to production capability.**
   b) **Application oriented Research, Design and Development (R,D&D) having production potential.**
   c) Basic R&D

5. **Objective of the Project:**

Broad Objectives

This project seeks to identify applications which will be of use to common people residing in say rural areas, small towns, or urban areas, power them through technology, and create local content for the application through the involvement of students and other users. Some of these applications will be enabled to work on cell phones so that they can reach large number of users. The setting up of the applications and portals would be through desktops, however. The local content (which would make many of the applications useful in the first place) would be through mass effort of crowd-sourcing/ community sourcing. To ensure that a good amount of content gets created, the effort will be linked to students in schools.

Objectives

1. To select, design and implement suitable applications which will be relevant to a community in a rural or urban area.

2. To power the applications with local content using a variety of community sourcing experiments on a pilot scale. The community sourcing would be of at least two types: school sourcing and village sourcing.

3. To design tools and services for mobile phone based web applications so as to provide access to a large number of users.

4. To undertake a social study of relationships and how the pilot level experiments make use of them, and in turn affect them, etc.

Thus, the project will experiment with new interactional forms which are expected to work in community sourcing contexts. Design human interaction tools that enable idea sharing, non-hierarchical decision making and proper utilization of the community web.

6. **Brief outline of the project with specific technology fall-outs:**

This project seeks to develop technology platform and implement applications which will permit a community web to be established. Tools will also be developed which will permit community sourcing of local content.

There will be three classes of community sourcing applications:

(A) **Family and Community Web**: To facilitate family well-being and to promote collectivity among communities through the use of information exchange and interaction.
(B) **Cultural and Geographical Resource Web**: To provide for sonic, visual and textual resource sourcing and maintenance.
(C) **Professions and Enterprise Web**: To enable peer group formations among self-employed professionals and to enhance their economic viability through information management.

Experiments will be done at the pilot level based on the success of which major scaling up can be done in subsequent project.

7. **Expected outcome in physical terms (as applicable)**

a) **Specifications of subsystem/system (as applicable)**

A major challenge in this project is the identification of applications in the rural community. Once this is done, suitable systems and sub-systems would be developed.

b) **Nature of documents for technology transfer**

The technology development will include not just software with proper engineering but also detailed documents to support technology transfer.

c) **Manpower trained**
   i) **Level of training**
   ii) **Nos. (industry/outside R&D/Internal)**

At the pilot level hands on training will be given on the tools for sourcing of content.

The people trained would be teachers and students in schools in the pilot area.

8. **Agency with which commercial link up is (Details may be given as applicable) established/ proposed.**

Not Applicable

9. **Likely End User(s)**

Several Villages in 3 regions located in 3 parts of the country.

10. **Revenue outcome proposed**:

There would be no revenue. Expenditure is listed under budget.

11. **Duration of Project**: 2 Years

12. **Year-wise break-up of physical achievements with specific intermediate milestones (in terms of aims and objectives)**

Within 3 months
   a) Finalize NGOs who would help in running the pilots.
   b) Put the initial central team together including identification or selection of research staff.

Within 6 months
   c) Teams at pilot locations to be in place.
   d) Workshop on selection of user applications and local content for pilots.
   e) Initial survey on community needs.

Within 9 Months
   f) Workshop with teachers on local sourcing of content.

g) Initial technology tools for local content creation (alpha-version) to be ready.
h) Training of teachers on tools for content creation and integrating them with courses in their education.

Within 12 months

i) First set of application launched with initial technology and some local content.
j) Social study completed on how web can help in fulfilling community needs and help village professionals.

Within 18 months

k) A set of half a dozen applications launched catering to schools and villages.

With in 24 months

l) Impact study of applications completed.
m) Workshop on impact study and future directions held.

13. **Name of other organisations jointly participating in the project (including organisation abroad)**

Not Applicable at this time

Signature of Chief Investigator
Designation
Date

Signature of
Head of the Institution/Organisation
Designation
Date

Additional Information Required

1. Wherever applicable, Under S.No.14, share of the industry, collaborating agency, any other assistance and DIT's support required in the total cost of the Project may be provided under various budget heads.
2. Brief history of the electronics company including products being made, capacities, related collaborators, achievements, capabilities etc. may be provided (including recent annual reports and company brochure)
3. Please indicate recent major achievements of in-house R&D Unit of the electronics company in development of new products/processes, technology export, patent taken etc. and whether in-house R&D unit of the firm is recognised by DSIR.
4. Any other information in support of the proposal.

DETAILS OF THE PROPOSAL

PART 1: BACKGROUND INFORMATION

1. **Title of Project**:- Socionity: Web With Community Sourcing of Local Content

2. **Chief Investigator:** Prof. Navjyoti Singh
Professor
Centre for Exact Humanities
International Institute of Information Technology
Gachibowli, Hyderabad 500 032
Andhra Pradesh, INDIA
Phone: +91-40-6653 1403, 6653 1302
Email: navjyoti@iiit.ac.in

3. **Other Investigators of the Project with their designations**

Co-Investigator: Dr. Rajeev Sangal
Professor & Director
Language Technologies Research Centre
International Institute of Information Technology
Gachibowli, Hyderabad 500 032
Andhra Pradesh, INDIA
Phone:+91-40-23001412
Email: sangal@iiit.ac.in

4. **Brief Bio-data of Chief Investigator and other Investigators (including publications/patents) (Please attach separate sheets)**

Enclosed

5. **Competence of Investigator in Project Area (Including Industry interaction/Technology transfer)**

a. Chief Investigator has done deep work on understanding Indian society and culture and how professionals can be supported.
b. Co-investigator has worked on NLP and developing language technology for Indian languages.

6. **Other Commitments of the Chief Investigator and Co-Investigators (including lectures, research projects responsibilities etc.) Indicate the percentage of time the Chief Investigator and Co-Investigator would devote to the project.**

Chief Investigator: 20%

Co-Investigator: 10%

7. **Details on each of the ongoing/completed projects with the Chief Investigator/Co-Investigator/R&D Team**

A.

i) Project Title: Indian Language to Indian Language Machine Translation System (ILMT)

ii) Funding Agency (or Internal funding): TDIL, DIT

iii) Brief Project Summary: Indian Language to Indian Language Machine Translation System (ILMT Project) is a bi-directional machine translation system, developed for nine Indian language pairs (18 Systems) by a Consortium of 11 Institutions with IIIT Hyderabad as Consortium Leader. This consortium project is funded by Department of Information Technology, Ministry of Communication and Information Technology.

iv) Technical Status vis-a-vis objectives :

- Develop machine translation (MT) systems from one Indian language to another Indian language. The chosen language pairs are listed below: The systems will be bi-directional and will work in general domain.
- We also plan to deliver domain specific versions of the MT system running in real life. The domains are: Tourism plus one other domain to be decided in the first 3 months of the project.
- It will also lead to the development of basic tools and lexical resources for Indian languages, such as POS taggers, chunkers, morph analysers, bidirectional bilingual dictionaries, annotated corpora, etc.

v) Financial Status: Total Project= 701.45 Lakhs, expenditure= 694.24 Lakhs)

vi) Duration and year of initiation: 3 years, August,2006

vii) Expected date of completion: Completed successfully on 30th April, 2010

B.

i) Project Title: Sanskrit-Hindi Machine Translation System (SHMT)

ii) Funding Agency (or Internal funding): TDIL, DIT

iii) Brief Project Summary: The objective of this project is to develop Sanskrit computational tools and use them to develop machine translation technology from Sanskrit to Hindi.

iv) Technical Status vis-a-vis objectives

- Develop machine translation (MT) systems from Sanskrit to Hindi.
- Deliver domain specific versions of the MT system. The domains are: Children stories and one scientific domain such as Ayurveda.
- Development of basic tools and lexical resources for Sanskrit, such as POS taggers, morph analysers, Sanskrit-Hindi bilingual dictionaries, annotated corpora, sandhi splitters for Sanskrit.
- Develop advance theories for sentence parsing for Indian languages by developing full fledged sentence parsers and treebank for Sanskrit. (Note: This will contribute to raising quality of many different applications for Indian languages such as Machine Translation, information search and extraction, etc.)

v) Financial Status (Total=53.73 Lakhs, expenditure=38.22 Lakhs)

vi) Duration and year of initiation: 3 Years, March 2008

vii) Expected date of completion: March 2011

C.

i) Project Title: English to Indian language Machine Translation System (EILMT)

ii) Funding Agency (or Internal funding): TDIL, DI

iii) Brief Project Summary: Develop machine translation (MT) systems from one English to Indian language in chosen language pairs – English – Hindi, English – Bengali, English – Marathi, English – Oriya, English – Urdu and English – Tamil.

iv) Technical Status vis-a-vis objectives

To develop basic tools and lexical resources such as Input Format Extractor, Morph analyser, Named Entity Identifier, POS tagger, Word Sense Disambiguation, Example Based Machine Translation, Statistical Machine Translation, TAG Based MT (Parser and Generator), Anal-Gen Based MT (Parser and Generator), Semantic Feature on TAG Trees, Post Processing Tools, Linguistic Resource Management Tools.

v) Financial Status (Total Project: 57.33, expenditure: 56.39)
vi) Duration and year of initiation: 3 Years, August 2006
vii) Expected date of completion: June 2010

8. **Brief summary of other project proposals (submitted by any of the Investigators) awaiting consideration of Media Lab Asia/DIT and other funding agencies like DST, DRDO, DSIR, MHRD, ICICI, IDBI etc.**

Nil

9. **Infrastructure and other facilities available at the institute for undertaking this project.**

a) List of major equipment along with model numbers, specifications etc.

Intranet infrastructure

b) Existing manpower and other personnel with names available for the project on full-time basis.

1. Mr. Harsh Satya
2. Mr. Devansh Mittal
3. Mr. Rajesh Tavva

10. **Expensive Equipment /facilities available elsewhere which could be made use of for the project.**

Nil

11. **Details of collaborating agencies (As this would vary from project to project, necessary details may be given as appropriate)**

Not applicable

12. **Additional information, if any.**

Nil

PART II : TECHNICAL INFORMATION

Title:- Socionity: Web with Community Sourcing of Local Content

1. Aim and Scope of the project (in terms of specific physical achievement)

Background

Information Technology has impacted life in India in a variety of ways, from train reservation to utilities and banks, and from e-governance to education. In many of these applications, the user has benefitted through a service provider using IT. There has been relatively little involvement of the user interacting with IT directly, or in contributing to development of IT services.

There have been several reasons for this lack of direct involvement as a direct user or contributor. The underlying reasons range from lack of access to computers, existence of language barrier, or lack of suitable user operable tools.

A qualitative change has come about today in situation pertaining to each of the underlying reasons. There is unprecedented proliferation of access devices today in the form of cell phones. Many standards together with enabling technology are in place today. For example, standards pertaining to Indian languages are in place, although there is the issue of learning curve for keyboards. Many user operable tools are available beginning from Wiki technology. What is needed is to put these elements together in applications, with the right technology and content and suitable social organization. A large number of people are capable of becoming direct users of, as well as contributors to, the IT enabled network, if the right applications are available.

At the social and human scale, there is a feeling that unless the electronic network and applications link with the human relationship in their full generality, they will not be able to play a fuller role in society and its development.

There is a need therefore to understand the role of physical community with respect to the internet. Applications would need to be built around family and community, which the internet and the application would naturally make use of. Thus, it is necessary to understand the society together with its family, social and economic relationships, and the ways in which the electronic network links and aids a rich human network. Socionity stands for society & community support by electronic network.

Broad Objectives

This project seeks to identify applications which will be of use to common people residing in say rural areas, small towns, or urban areas, power them through technology, and create local content for the application through the involvement of students and other users. Some of these applications will be enabled to work on cell phones so that they can reach large number of users. The setting up of the applications and portals would be through desktops, however. The local content (which would make many of the applications useful in the first place) would be through mass effort of crowd-sourcing/ community sourcing. To ensure that a good amount of content gets created, the effort will be linked to students in schools.

Objectives

1. To select, design and implement suitable applications which will be relevant to a community in a rural or urban area.

2. To power the applications with local content using a variety of community sourcing experiments on a pilot scale. The community sourcing would be of at least two types: school sourcing and village sourcing.

3. To design tools and services for mobile phone based web applications so as to provide access to a large number of users.

4. To undertake a social study of relationships and how the pilot level experiments make use of them, and in turn affect them, etc.

Thus, the project will experiment with new interactional forms which are expected to work in community sourcing contexts. Design human interaction tools that enable idea sharing, non-hierarchical decision making and proper utilization of the community web.

2. Detailed description of the Project

2.1 Introduction

Web is fast becoming ubiquitous. However, digital divide in India is a glaring reality as well. Presence of Indian people and culture on web is significantly low. The project intends to reverse this situation forever and bring web to the doorsteps of ordinary Indian people and to appreciably enhance Indian content on the web. This would be done by conducting wide variety of experiments on community sourcing.

In ordinary Indian people, though the literacy is low, the knowledge content is reasonably high and multifarious. Such knowledge about culture, livelihood, nature, places, profession etc. is orally transacted. Acts of cooperation, coordination and collectivity are orally transacted. We need to capture this information to harness community sourcing of this knowledge and activities. Mobile is a computer in the hands of Indian people, which is under utilized. Community web which also allow mobile based oral access could energize ordinary people to cultivate their participation in the web in a way that enhances their quality of life.

Web based enhancement of the quality of life must touch fundamentals of life. It should make elementary social and economic sense besides expanding horizons of life through global connectivity. It should further family well-being, community activities and professional activities of ordinary people. One can also look beyond client-server paradigm towards peer-to-peer paradigm to conceive radical jump in the web participation of ordinary people.

The information deficit is one of the causes of depressed/defeated life. The sense of being left out is disabling. This is so particularly with self-employed people, who demographically form large section of Indian society. Such people constitute bulk of rural India, artisanal India and service India, the underbelly of Indian society. Appropriate information applications can regenerate viability of their self-employment. Profitable peer-to-peer enterprises of self-employed can be constructed since community web would be geography aware. Professional web communities can be physical communities as well.

Apart from economic sense, community web will make social sense for underprivileged. community sourcing for family web, for community web, for village and town web is quite plausible. Community sourcing for cultural web, for educational web, for natural resources web etc. can be experimented with. Similarly, for grass-root democracy, transparency of governance, web based access to governmental

schemes, development requirements etc. In fact community sourcing opportunities are incredibly many and calls for wide variety of experimentation as the project plans to do.

Youth from underprivileged section of society in collaboration with socially aware agencies/institutions is to be targeted for content sourcing and information-based local enterprises. Village and community schools can be hub for family, community, resources, educational and developmental information sourcing. They can be the proactive voice on the web. Being meaningfully connected and also being content generator can be empowering. We propose to conduct variety of experiments to locate appropriate segment in society that can fuel content generation on the oral-cum-traditional web. Full force of digital communication and web technology along with innovative applications has to be brought to the youth who would fuel community sourcing.

Community web which is geography aware and family oriented can bridge digital divide and radically enrich the web in terms of content. Such a web will of course be integrated with relevant web resources that are already present.

The web approach we take is (a) Community-centered (family centric), (b) Geography aware, (c) Acting locally, and (d) Thinking globally. The web methodology we adopt is (a) community sourcing, (b) Linking with education (student man power with curriculum), and (c) Oral web (text and multimedia enabled).

2.2 Community Sourcing Experiments

Larger part of the project is community sourcing application design. There are three classes of community sourcing applications:

(A) **Family and Community Web**: To facilitate family well-being and to promote collectivity among communities through the use of information exchange and interaction.
(B) **Cultural and Geographical Resource Web**: To provide for sonic, visual and textual resource sourcing and maintenance.
(C) **Professions and Enterprise Web**: To enable peer group formations among self-employed professionals and to enhance their economic viability through information management.

Primary focus in the project will be on (A) and (B), but at least one example would be attempted from (C). Some example applications are given below.

2.2.1. Family/Community Web:

Purpose: To facilitate family well-being and to promote collectivity among communities through the use of information exchange and interaction.

Web served by a Village Server: Possibly kept and maintained at the village school by teachers/students with help from NGO.

Culture

1. Personae wiki
2. Folk songs
3. Village stories
4. Village itihaas
    a. Including floods in village

b. Earthquakes etc.
c. Great social events

Nature related

1. Bird watchers
    a. Bird count
2. Animals watchers
3. Tree watchers
4. Latest trends – mapping size, water and quality
5. River flow and water quality
6. Maximum temperature, rainfall, humidity
7. Wind direction – daily
8. Tubewell water level and water quality

Connectivity

1. Nearest airport, railway station, bus stop
2. River way – distance etc.

Communication and Media

1. Village newspaper/mandal or block newspaper
2. SMS-mailing groups – village, family, block etc.
3. E-mailing groups – village, family, blocks etc.

Local Language and dialect

1. Dictionary of local words (say 1K to 5K words)
2. Grammar and constructions

2.2.2 Cultural/Geographical Resource Web

Purpose: To provide for Sonic, Visual and Textual Resource Sourcing and Maintenance.

Central servers of the project maintained at CEH IIIT-H

Tasks:

1. Festivity Resources
2. Heritage Resources (Monuments, Visuals, Public Buildings)
3. Stories, Literature, Mythology Resources
4. Persona Resources (to begin with Gandhi, Tagore, Aurobindo, Ambedkar collective works)
5. Geography Resources (Cities, Towns, Villages etc.)
6. Tourism/Hospitality Resources
7. Textual Heritage Resources (oral and written texts)
8. History Resources
9. Performing Arts Resources
10. Fine Arts Resources
11. Languages Resources
12. Cultural Zones
13. Natural Resources

2.2.3 Professions/Enterprise Web

Purpose: To enable peer group formations among self-employed professionals and to enhance economic viability of professions through information management.

Mobile served by a Professional Community Server: Possibly kept and maintained at the community school or community organizations

Artisanal Professions and Communities

1. Artisanal Community Wiki
2. Products data
3. Organizations and Professional Society data
4. Economic/Market data
5. Forums

Artistic Professions and Communities

1. Artist community Wiki
2. Product previews and recordings
3. Event organizer
4. Enterprise portal
5. Forums

Services Professionals and Communities

1. Service Community Wiki
2. Classes of services
3. Service resources in localities and services' organizer
4. Forums
    - Animal Husbandry and Fishery
    - Agriculture, Horticulture and Forestry
    - Village Industry (Brick, Stone Cutters, Sericulture etc.)
    - Public Amenities Maintenance Professionals
    - Labor Market
    - Health Professionals
    - Transportation Professionals
    - Institutional Infrastructure (Banks, Post, School, Dispensary, Community Hall)
    - Professional Markets
    - Sales Promotion and Management
    - Employment Opportunities
    - Entrepreneur Opportunities
    - Government offices, services and promotion schemes

2.3 **Methodology & Strategy**

We will work with several pilots, running in different regions. For example, pilots could run in the vicinity of Hyderabad, Chennai, Tirupati, Raipur, Kanpur, Delhi and Mussorie-Kempty. Each pilot will run in a number of villages in its geographic area. Three pilots would be chosen out of the above.

An exercise would be carried out to select suitable applications. The applications would be so chosen that they are of relevance to the community in the area. They should have a strong component of local content which can be contributed by the children in the schools in that area, as well as others. Primarily, government schools or NGO run schools would be chosen in each area.

For each group, we would also select an NGO with whom we would work. There would also be an attempt to involve the local community in the development and use of the application, including sourcing of content.

Humanities and Social Sciences research group at IIIT-H would be involved in studying the community so that suitable applications can be chosen. It would help in the design of the application as well. The research group would also study the impact of the application on the community and the role internet plays in the physical community. There are other areas of study as well, wherein the web allows for family presence rather than only individual presence in the cyberspace.

Technology development for the application would be carried out at IIIT Hyderabad. The students who would develop the application would work in close collaboration with the end users. The applications would be enabled in the local languages and dialects.

The project would be carried out over a 2-year period.

2.4 Project Plan

In the first three months of the project, NGOs will be identified who can be part of the project and help in running the pilots. The NGO's programs will be studied and analyzed to get the relevant information, this information will be used to select the NGO. Out of the 7 regions mentioned in the proposal, 3 regions with an NGO in each will be chosen.

After the selection of the NGOs, their team would have to be put together by them. The initial central team will be put in place at Hyderabad by recruiting the research staff (including humanities staff/experts) and programming staff.

A workshop will be conducted to identify and select the end user applications and local content for the pilots through brainstorming sessions, and discussions. This would benefit from the rich experience of the NGO leaders.

An initial survey would also be conducted in the villages to identify community needs and what steps would strengthen the community. The results of the initial survey would be analyzed to see how the selected applications would fit with the community. Particular emphasis would be given to see what kind of local content would need to be created, and how it would fit in with the school curriculum.

Initial technology tools for the local content will be prepared with rapid prototyping. Training will be given to the teachers on the local sourcing of the content using technology tools. Stand-alone software would be installed at all the pilot locations, so there is no dependence on the web while creating the content.

Within 12 months, first set of application will be launched with initial technology, and some local content would be created by the teacher, local staff and students. Social study done during the entire year will answer the questions as to how the web can help in fulfilling community and village needs.

Large amount of content would be created over the next 12 months. Experience of applications in the pilots would also be studied to know the similarities and differences across villages and pilots. This will help to enrich knowledge about the applications. Later, the users and school children will themselves identify, suggest, and even implement applications.

By the end of two years, several real applications will be running and valuable local content would be created. The impact study of the pilots would be completed.

2.5 References:

- James Surowiecki, *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*, London: Little Brown, 2004.
- Jeff Howe, *Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business*, CA: Three Rivers Press, 2008.
- Don Tapscott, Anthony D. Williams, *Wikinomics: How Mass Collaboration Changes Everything,* Atlantic Books,2008.

3. **Need, forecast and urgency for the technology proposed to be developed with justification such as importance of know-how, import substitution role, pay off w.r.t. purchase of know-how or development of technology competitiveness, technology exports, international alliances possibilities etc**.

   As discussed under point 1 in background, this project proposes to explore and identify useful applications based on the internet which will also connect with the local community and school children in sourcing of local content.

4. **Specific manner in which know-how generated here is envisaged to be translated into production, details regarding**

   a) The end product (with specifications to be attained etc.)

   b) **Availability of pilot production facility in the organization**

      Not applicable

5. **a) Name of production agencies willing to productionise/use and market surveys if any made by them regarding demand for the product**

   b) **Alternative production/user agencies.**

      Not applicable

6. **Period required for completing the project:** Two Years

7. **Details of work already done by present investigators/ R&D team in this or other areas**

   a) Successfully completed on schedule :

      Indian Language to Indian Language Machine Translation System (ILMT): ILMT is a bi-directional machine translation system, developed for nine Indian language pairs (18 Systems) by

a Consortium of 11 Institutions with IIIT Hyderabad as Consortium Leader. This consortium project is funded by Department of Information Technology, Ministry of Communication and Information Technology.

English to Indian language Machine Translation System (EILMT): EILMT is a bidirectional system developed for following language pairs – English – Hindi, English – Bengali, English – Marathi, English – Oriya, English – Urdu and English – Tamil. This consortium project is funded by Department of Information Technology, Ministry of Communication and Information Technology.

EBMT Lexical resource creation: Two MT systems have been developed which test and validate the language databases:
1.Shiva MT system (in collaboration with CMU USA, IISc Babglore. Shiva is an example based MT system
2. Shakti MT system: Shakti follows a hybrid approach. It uses the linguistic knowledge as well as machine learning techniques. Shakti MT system is implemented for English-Hindi, English-Telugu and English-Marathi

b) **Currently in progress:**

Indian Language to Indian Language Machine Translation System Phase-II: (ILMT- PII)
The objectives of the project is to:

1. Enhance the Sampark Indian Language Machine Translation System (ILMT) for real time, on line deployment over the web for general use.
2. Identify specific user organization and integrate two of the Sampark MT System with their application in their workflow.

To enhance accuracy and performance (or speed) of the current IL machine translation systems a number of developments are needed. The systems should also be able to meet the requirements of real-life systems, with a given level of translation accuracy, and they should be capable of further improvement. For achieving this some users would be identified whose requirements would be used as inputs for further enhancing the systems to the expected level of comprehensibility and accuracy.

Dashboard: The Dashboard Development Environment is based on Blackboard architecture which facilitates a set of components (can be heterogeneous as well) collaborate among themselves, through operating on a common memory (blackboard), by applying their intelligence to solve a complex problem. Apart from implementing the features of blackboard architecture, we also plan to develop many more additional tools to enhance it as a development environment for a family of NLP Applications, such as Machine Translation System, Speech to Speech Translation System, Information Extraction System, etc. The proposed Dashboard Development Environment would help: development and testing of component in isolation, integration and testing of a set of heterogeneous components, building and testing of complete system, facilitate the training and tuning of the system, and regression testing as well. It would have extremely user-friendly graphical user interface to operate with different access rights

c) **Abandoned:** Nil

d) **Industry interaction/know-how transferred**

8. **Summary of similar work being done elsewhere in the country**

The projects below have focused on specific areas such as agriculture successfully. They have not focused on community building and sourcing of local content by local people, which is the goal of this project. Hence this project would work on a complementary aspect.

E-sagu: Since 2004, by exploiting ICTs, efforts are being made to develop eSagu system to enable Indian farmers to improve the farm productivity under Indian situtaion. The eSagu system is an ICT-based personalized agro-advisory system. ("Sagu" means cultivation in Telugu language.). In this system, high quality personalized (farm-specific) agro-expert advice in a timely manner to each farm at the farmer's door-steps. In e-Sagu, rather than visiting the crop in person, the agricultural scientist delivers the expert advice by getting the crop status in the form of digital photographs and other information. The expert advice is generated by agricultural experts based on the latest information about the crop situation received in the form of both digital photographs and corresponding text. The advice is provided on regular basis from sowing to harvesting which reduces the cost of cultivation and increases the farm productivity as well as quality of agri-commodities. In eSagu, the developments in ICT such as (database, Internet, and digital photography) are extended to improve the performance of agricultural extension services. The impact results show that the farmers are getting additional benfit of Rs 3,820/- by reducing fertilizer and pesticide inputs, and getting extra yield. The eSagu project has won several awards.

The institute has gained valuable experience in dealing with a project running in villages with local people supported by the central team.

e- choupal: ITC's strategic intent is to develop e-Choupal as a significant two-way multidimensional delivery channel, efficiently carrying goods and services out of and into rural India. By progressively linking the digital infrastructure to a physical network of rural business hubs and agro- extension services, ITC is transforming the way farmers do business, and the way rural markets work.

The network of 6,500 e-Choupal centres spread across 40,000 villages has emerged as the gateway of an expanding spectrum of commodities leaving farms – wheat, rice, pulses, soya, maize, spices, coffee, aqua-products. The reverse flow carries FMCG, durables, automotives and banking services back to villages.

e-choupal does not relate to existing community in rural areas or the current system of procurement of produce for obvious reasons.

9. **Information regarding specific intermediate milestones (year-wise)**

Milestones:

Within 3 months

n) Finalize NGOs who would help in running the pilots.
o) Put the initial central team together including identification or selection of research staff.

Within 6 months

p) Teams at pilot locations to be in place.
q) Workshop on selection of user applications and local content for pilots.
r) Initial survey on community needs.

Within 9 Months

s) Workshop with teachers on local sourcing of content.
t) Initial technology tools for local content creation (alpha-version) to be ready.
u) Training of teachers on tools for content creation and integrating them with courses in their education.

Within 12 months

v) First set of application launched with initial technology and some local content.
w) Social study completed on how web can help in fulfilling community needs and help village professionals.

Within 18 months

x) A set of half a dozen applications launched catering to schools and villages.

With in 24 months

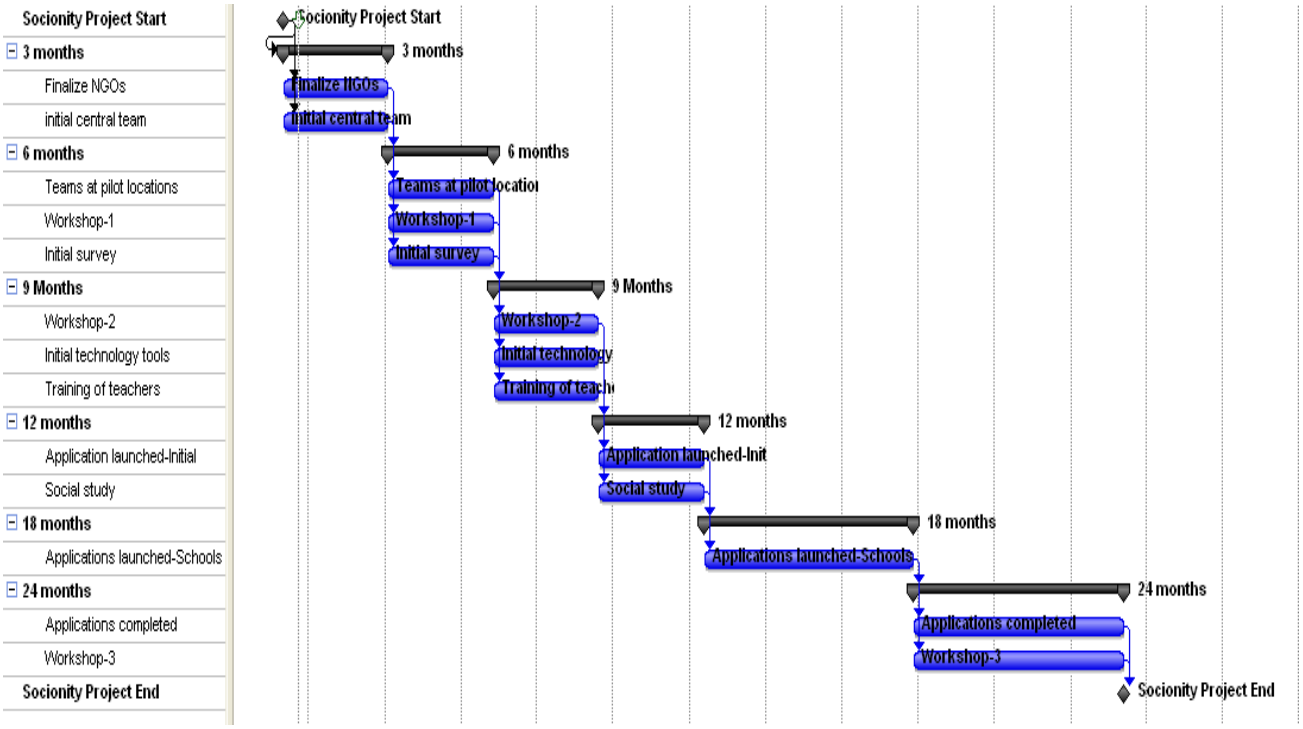
y) Impact study of applications completed.
z) Workshop on impact study and future directions held.

10. **a) Specific problems, hold-ups and difficulties foreseen in the implementation of the project.**

No major difficulties foreseen. Initial difficulties of setting up experiments in villages will be solved with the help of local NGOs working closely with schools.

b) **If the answer is not Nil to 10(a), how does Chief Investigator propose to overcome them?**

11. **Detailed PERT/BAR Chart (Separate Sheet)**



12. **Details of possible alternative arrangements if the Chief Investigator leaves institution or is unable for any other reason to continue on this project.**

    Other investigators will carry the project forward.

13. **Name of other organisations in India or Abroad jointly participating in this effort, extent of their involvement, specific division of responsibility, accountability etc.**

    As mentioned, the responsibility of running the pilot would fall on the selected NGOs. The NGO because of its past history of work and rural linkages would be able to carry the pilot forward rapidly.

    The possible NGOs are:

    a) **SIDH, Society for Integrated Development for Himalayas, Mussoorie-Kempty: S**IDH has been working since 1989 in the field of education. It is led by Shri. Pawan Gupta, BTech from IIT Delhi and his wife. It is running several schools and has done fundamental work in the pedagogy as well as on content of education.

    b) **Manaviya Shiksha Sanskar Sansthan (MS3), Mandana, Kanpur: M**S3 has been working on human values and has been running workshops for teachers, local villagers, industry people etc. It is led by Shri Ganesh Bagaria (BTech and MTech IIT Kanpur in 1982 and 1985 respectively) and his wife.

    It has had a tremendous impact on people from different walks of life. MS3 has wide contacts with villagers, and a large sphere of influence. They will be able to connect with local schools and teachers. They also have a band of dedicated IIT Kanpur students who work with them closely.

c) **Center for ICT in Rural Development (CICTRD), village Venkatrampuram, Dist. Tirupati :** CICTRD works on web-based free-and-open source software for rural ommunities. It is run by Nagesh Kolagani (BTech from IIT-Madras in 1988 and MS (Computer Vision/ Robotics) from Hawaii, USA in 1992) and his wife Aparna (B.Tech. Computer Science) from their village Venkatrampuram. Before settling in the village 14 years ago they worked as software engineer during 1992-95 in California and Infosys respectively. They are involved in improving and promoting village life in various ways.

d) **Samanvaya Knowledge Trust (SKT), Chennai:** SKT is an NGO dedicated to the pursuit of true and natural learning through people's initiative It is led by Ram Subramanian and his wife Ramaa. It has been involved with Gandhian educational program of Nai Taliim and with multiversity education model. It is also involved with organic farming experiments in Chennai region with several partners.

e) **Abhyudaya Sansthan, Achhauti, Durg-Raipur:** Abhyudaya Sansthan started as an experiment on community living based on agriculture. It has a large education program on Human Values. It is engaged in a massive program of teacher training for government schools in Chhattisgarh on human values.

14. **List the personnel already working in the organisation who would be transferred to work full time on this project.**

    a) Mr. Harsh Satya
    b) Mr. Devansh Mittal
    c) Mr. Rajesh Tavva

    Their theses work is in similar area and would relate to this project. This will allow them to devote substantial amount of time on this project.

15. **Name of experts whom the Chief Investigator would invite to join the project team as full time/part time member.**

    Experts at different locations will not join as full time salaried employees. They will continue to work from their NGO locations.

PART III - FINANCIAL DETAILS

Table I: Yearly Break-up

Year I : Break-up							
S.No.	**Head**	**Local expenses**	**Foreign Exchange (FE)**	**Duty**	**Total**	**Part of 6 to be borne by participating/other organizations**	**Amount payable by DIT**
1	Capital Equipment	7.80	0	0	7.80	0	7.80
2	Consumable Stores	1.00	0	0	1.00	0	1.00
3	Manpower	21.30	0	0	21.30	0	21.30
4	Travel/Training						
	Travel	2.00	**0**	**0**	2.00	**0**	**2.00**
	Workshop	0.75	**0**	**0**	0.75	**0**	**0.75**
	per person *	2.25	**0**	**0**	2.25	**0**	**2.25**
5	Contingencies	1.00	**0**	**0**	1.00	**0**	**1.00**
6	Overheads	5.42			5.42		5.42
	Total	**41.52**	**0**	**0**	**41.52**	**0**	**41.52**

Year II : Break-up							
S.No.	**Head**	**Local**	**Foreign**	**Duty**	**Total**	**Part of 6 to be**	**Amount**

		expenses	Exchange (FE)			borne by participating/other organizations	payable by DIT
1	Capital Equipment	0.00	0	0	0.00	0	0.00
2	Consumable Stores	1.00	0	0	1.00	0	1.00
3	Manpower	42.60	0	0	42.60	0	42.60
4	Travel/Training						
	Travel	**2.00**	**0**	**0**	2.00	**0**	**2.00**
	Workshop	**0.75**	**0**	**0**	0.75	**0**	**0.75**
	per person *	2.25	**0**	**0**	2.25	**0**	**2.25**
5	Contingencies	1.00	**0**	**0**	1.00	**0**	**1.00**
6	Overheads	7.44			5.42		5.42
	Total	**57.04**	**0**	**0**	**57.04**	**0**	**57.04**

*** Cost of stay 200 days in villagae @ Rs.100 p.d = .21 per person per year**

Total Budget Outlay

S No	Head	Local expenses	Foreign Exchange (FE)	Duty	Total	Part of 6 to be borne by participating/ other organizations	Amount payable by DIT
1	Capital Equipment	7.80	0.00	0.00	7.80	0.00	7.80
2	Consumable Stores	2.00	0.00	0.00	2.00	0.00	2.00
3	Manpower	63.90	0.00	0.00	63.90	0.00	63.90
4	Travel/ Training	10.00	0.00	0.00	10.00	0.00	10.00
5	Contingencies	2.00	0.00	0.00	2.00	0.00	2.00
6	Overheads	12.86	0.00	0.00	12.86	0.00	12.86
7	**Total**	**98.56**	**0.00**	**0.00**	**98.56**	**0.00**	**98.56**

Table II: Subsystem wise Break-up

Subsystem wise Break-up

S No	Item Description	Local	Foreign Exchange	Duty	Freight	Total
1	26 Personal computers @ 0.3 each (Per pilot 8 PC, Mini pilot 4 PC and Central team 6 PC)	**7.80**	**0**	**0**	**0**	**7.80**
	Total	**7.80**	**0**	**0**	**0**	**7.80**

TABLE-III: MANPOWER DETAILS

Manpower details per Pilot							
S.No.	**Designation of post**	**Yearly salary in lakhs**	**1st Year**		**2nd Year**		**Total**
			No.of Posts (Budged for 6 months)	**Total Expenditure**	**No.of posts**	**Total Expenditure**	**(Rs. lakhs)**
1	Sr Research Engineer	6	0.5	3	1	6	9
2	Research staff	3	1	3	2	6	9
3	Local Staff	1.2	1	1.2	2	2.4	3.6
4	**Total**		**2.5**	**7.2**	**5**	**14.4**	**21.6**

Manpower details for Mini Pilot							
S.No.	**Designation of post**	**Yearly salary in lakhs**	**1st Year**		**2nd Year**		**Total**
			No.of Posts (Budged for 6 months)	**Total Expenditure**	**No.of posts**	**Total Expenditure**	**(Rs. lakhs)**
1	**2**	**3**	**4**		**5**		**6**
1	Research Engineer	2.4	0.5	1.2	1	2.4	3.6
2	Local Staff	1.2	1	1.2	2	2.4	3.6
3	**Total**		**1.5**	**2.4**	**3**	**4.8**	**7.2**

Manpower details of Central Team							
S.No.	**Designation of post**	**Yearly salary in lakhs**	**1st Year**		**2nd Year**		**Total**
			No.of Posts (Budged for 6 months)	**Total Expenditure**	**No.of posts**	**Total Expenditure**	**(Rs. lakhs)**
1	Jr CS	1.2	2.5	3	5	6	9
2	Admin	1.5	1	1.5	2	3	4.5
3	**Total**		**3.5**	**4.5**	**7**	**9**	**13.5**

Total Manpower details (includes 2 pilots, 1 mini pilot and central team)

S.No.	**Designation of post**	**Yearly salary in lakhs**	**1st Year**		**2nd Year**		**Total**
			No.of Posts (Budged for 6 months	**Total Expenditure**	**No.of posts**	**Total Expenditure**	**(Rs. lakhs)**
1	Sr Research Engineer	6	1	6	2	12	18
2	Research Engineer	2.4	0.5	1.2	1	2.4	3.6
3	Research staff	3	2	6	4	12	18
4	Jr CS	1.2	2.5	3	5	6	9
5	Admin	1.5	1	1.5	2	3	4.5
6	Local Staff	1.2	3	3.6	6	7.2	10.8
7	**Total**			**21.3**	20	**42.6**	**63.9**

Part IV

Endorsement by the Head of the Institution

1. I have read the terms & conditions (including special terms & conditions for co-financing) governing the grant-in-aid and I agree to abide by them.

2. I certify that I have no objection to the submission of this research proposal "Socionity: Web with Community Sourcing of Local Content" for consideration by the Ministry of Information Technology

3. In case the project is approved, I undertake to make available facilities to carry it out, to arrange for the submission of periodic progress reports and other information that may be required by the Ministry of Information Technology and In general to ensure that the conditions attached to the award of such grant are fulfilled by my institution/organisation.

4. I certify that in case present chief investigator is not available for any reason to continue work on this project, the following persons will be available to carry it throughout to completion:

Sl.No.	**Name**	**Designation**
1.	Pradeep Kumar. R	Associate Professor
2.		

5. I certify that the facilities mentioned in the body of this report are available at my institution.

6. I certify that I shall ensure that accounts will be kept of the funds received and spent and made available on demand, as specified and required by the DIT.

7. I certify that I am the competent authority, the virtue of the administrative and financial powers vested in me by to undertake the above stated commitments on behalf of my institution.

Signature of the
Head of the Institution
Designation
Date: