**TEAM: The Anaphora resolution**

**Languages/Ideas!!**

POS Tagging - Brill Tagger - Hindi

Sanskrit - Very logical -Corpus?? Less ambiguity -> What kind of task?

Classification of documents into classes -|||- Naive bayes -

Hindi dialects? Dataset?

Morphology - ->

Stemmer.

Idea 1: Intelligent text question answering machine

1. Take in a corpus in Hindi, and draw inference out in a prolog based knowledge base
2. User asks questions in Hindi. Apply natural language processing techniques to parse and get relevant meaning out of question.
3. Apply AI techniques to get answer to question.
4. Display short answers to the questions.
5. Note: Query domain and types will be restricted.

Idea 2: Geographic Location Dissemination System

1. Questions asked by user in Hindi
2. Apply NLP techniques to get relevant info from the knowledge base
3. Display answer as a geographic data- eg. display on map?
4. Example,

Q- What are the neighbors of US?

A- Canada and Mexico

Idea 3: Improvement on Sanskrit language Stemmer + Word Segmentation

Some more details would be useful:

Idea 1:

1. How would inferences be drawn from the corpus?
2. How would meanings be extracted from the question, how would they be represented, and how would that be used to query the knowledge base? What techniques would be used to get the answers?

Idea 2: Same as above --

1. How would inferences be drawn from the corpus?

Corpus -> Wikipedia English - Geography

Crawler -> Extract ! (No NLP here)

We get a text finally for each country/location

Tokenize : then stemming

1. How would meanings be extracted from the question, how would they be represented, and how would that be used to query the knowledge base? What techniques would be used to get the answers?

how would the question be transformed into a geographical database query?

Idea 3: What is the starting point? What improvements would be made, and how can they be achieved?

Q1: All the ideas are fine in principle, but they need to be developed in more detail in order to see f they're viable.

Q2: Creating an application is great, but it shouldn't consume too much of the effort; the majority of the effort should be spent on language processing.

Q1. Are these ideas viable to work on? Please provide input on all of these ideas.

Q2. If we use any predefined NLP library, can we use other algorithmic techniques to create an application around it. Or is it essential to work upon NLP algorithms only?

Idea 4: <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=3&cad=rja&uact=8&ved=0ahUKEwjQtM6PpMbLAhVK6mMKHe9HAQ4QFggyMAI&url=http%3A%2F%2Fwww.rroij.com%2Fopen-access%2Fsimulation-of-opinion-mining-in-hindilanguage-based-on-natural-languageprocessing.pdf&usg=AFQjCNHYDjGPlblnmcupNFrGyjL9O0egBg&sig2=hgS36xSXOXnDkkkLwBilDw>

Graph representation of geographical data:  
<https://agile-online.org/Conference_Paper/CDs/agile_2004/papers/6-1-3_Karalopoulos.pdf>

Corpus other than in English:

<http://www.essex.ac.uk/linguistics/external/clmt/w3c/corpus_ling/content/corpora/list/index2.html>

Query system which supports geography

<http://start.csail.mit.edu/index.php>

<https://www.researchgate.net/publication/226232668_A_common_sense_geographic_knowledge_base_for_GIR>

Hindi Geography names list:

<http://unstats.un.org/unsd/geoinfo/UNGEGN/docs/5th-uncsgn-doc/e_conf_79_inf65.pdf>

Free GIS datasets:  
<http://gisgeography.com/best-free-gis-data-sources-raster-vector/>

Query answering system for Hindi

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.109.8228&rep=rep1&type=pdf>

Sentence structure parsing

<http://www.link.cs.cmu.edu/link/submit-sentence-4.html>

INDIC LIBRARY:

<http://anoopkunchukuttan.github.io/indic_nlp_library/>

POS Tagging in HINDI:

<http://airccj.org/CSCP/vol3/csit3639.pdf>

Subject-object-verb tagging (English):  
<http://stackoverflow.com/questions/18968457/is-there-a-reasonably-accurate-heuristic-for-detecting-the-subject-and-object-of>

HINDI TOOLKIT FOR HINDI POS:

<http://sivareddy.in/downloads#hindi_tools>

Hindi Shallow Parser:  
<http://ltrc.iiit.ac.in/analyzer/hindi/>

Stanford parser:

<http://nlp.stanford.edu:8080/parser/>

“If I understand correctly, the idea is to take a user question in Hindi and transform it into a query for a geographic database. This is definitely an NLP task, and the difficulty lies in the types of questions you'll be able to handle. For that you'd need a **reasonable corpus of geographical questions in Hindi (**Have to search for this) -- perhaps there's one available, or else you can collect one yourselves.

What type of parsing you need (if any) is not clear at this stage. You need to extract enough information from the Hindi sentence to fill out a database query. So the first thing to look at is what **kinds of queries** you intend to handle, and then see how you get the required information from the language input. Perhaps **entity extraction** and some **basic relation identification** can get you there, and perhaps you need more.

“

Things to Do:

Resources:

http://www.gkinhindi.net/2015/05/india-geography-questions.html

####################################

Setting up database with geographical data

####################################

data source: http://www.diva-gis.org/gdata

1. install postgre sql

2. install extension postGIS for postgre sql (required for spatial features)

3. bin location:

/Library/PostgreSQL/9.5/bin

4. add postgis extension:

run psql

at postgresql prompt:

CREATE EXTENSION postgis;

5. commands to import shp data:

#this creates a geo-spatial sql file that can be loaded into the db

/Library/PostgreSQL/9.5/bin/shp2pgsql data/IND\_water\_areas\_dcw.shp water\_areas postgres > water\_areas.sql

#this loads the data

/Library/PostgreSQL/9.5/bin/psql -d postgres -U postgres -p 5432 -f water\_areas.sql

NOTE: inland water bodies (rivers and lakes) for India has been loaded into the database at this point of time

SAMPLE QUERIES:

----------------------------

SELECT DISTINCT name from water\_areas;

SELECT gid, iso, country, name, hyc\_descri, f\_code\_des FROM water\_areas;

SELECT ST\_Area(geom) As sqft FROM water\_areas;

Query templates:

1. Neighboring states:

SELECT T3.N2, ST\_Touches(T3.G1, T3.G2) AS rval FROM (SELECT T1.name\_1 AS N1, T1.geom AS G1, T2.name\_1 AS N2, T2.geom AS G2 FROM administrative1 AS T1 CROSS JOIN administrative1 AS T2 where T1.name\_1 like 'Bihar') AS T3 where ST\_Touches(T3.G1, T3.G2) IS TRUE;

2. Distance between cities:

SELECT ST\_Distance(T3.G1, T3.G2) FROM (SELECT T1.name\_2 AS N1, T1.geom AS G1, T2.name\_2 AS N2, T2.geom AS G2 FROM administrative2 AS T1 CROSS JOIN administrative2 AS T2 where T1.name\_2 like 'Gaya' and T2.name\_2 like 'Bhojpur') AS T3;

3. What all states have a river flowing in them

SELECT DISTINCT T3.N2, ST\_Intersects(T3.G1, T3.G2) FROM (SELECT T1.nam AS N1, T1.geom AS G1, T2.name\_1 AS N2, T2.geom AS G2 FROM water\_lines AS T1 CROSS JOIN administrative1 AS T2 where T1.nam like 'MAHANADI') AS T3 where ST\_Intersects(T3.G1, T3.G2) IS TRUE;

**RAW CORPUS**

दिल्ली से कानपूर की दुरी कितनी है?

महारास्ट्र की छेत्रफल कितनी है?

राँची से सबसे पास कौन सी नदी है?

कौन से राज्य की जनसँख्या सबसे ज्यादा है?

भारत की सबसे दक्षिण शहर कौन सी है?

भारत में कितने राज्य हैं?

भारत में सबसे बड़ा झील कौन सा है?

गंगा कितने राज्यों में बहती है?

भारत में कितने राज्य समुन्द्र तट पे स्थित हैं?

बंगलौर से दिल्ली जाने वाली सड़क कितनी लम्बी है?

राज्यस्थान के कितनी पडोसी राज्य हैं?

मध्य प्रदेश के उत्तर में कौन सा राज्य है?

कश्मीर से कितनी नदियां प्रारंभ करती हैं?

भारत में किस राज्य की जनसँख्या सबसे कम है?

भारत के पांच सबसे छोटे राज्य कौन से हैं?

मुंबई से गुहाटी की दुरी कितनी है?

बिहार में कितने सहर हैं?

छत्तीसगढ़ से पूर्व कितने राज्य हैं?

भारत में सबसे ज्यादा हरियाली किस राज्य में है?

भारत में सबसे बंजर राज्य कौन सा है?

क्या दिल्ली बजनर राज्य है?

क्या बिहार समुद्र तट पर सिमित है?

बंगलौर किस राज्य में स्थित है?

क्या रांची और धनबाद एक ही राज्य में स्थित हैं?

कन्याकुमारी से कितने रेलरोड उत्पन्न होते हैं?

नर्मदा नदी किस राज्य से बहती है ?

बैंगलोर की ऊचाई समुद्री तल से कितनी है ?

गंगा नदी किस उत्पन्न होती है ?

लखनऊ किस राज्य में स्थित है ?

महाराष्ट्र की राजधानी क्या है ?

भारत की राजधानी क्या है ?

जयपुर किस राज्य की राजधानी है ?

किस राज्य का क्षेत्रफल सबसे ज्यादा है ?

बिहार का क्षेत्रफल क्या है ?

भारत का सबसे उत्तरी राज्य कौन सा है ?

कौन सा शहर भारत के दक्षिणी छोर पर स्थित है ?

सतलुज जिस राज्य से बहती है , उसका नाम क्या है ?

क्या एवेरेस्ट पहाड़ भारत में है ?

भारत की सबसे ऊची जगह कौन सी है ?

भारत के पूर्वी छोर पे कौन सा राज्य स्थित है ?

भारत के दक्षिण में कौन सा महासागर है ?

कौन से राज्य की जनसँख्या सबसे ज्यादा है ?

दिल्ली से आगरा की दूरी कितनी है ?

महाराष्ट्र से बिहार कितना दूर है ?

भारत में कितने केंद्र शासित प्रदेश है ?

भारत में कितने राज्य है ?

भारत में राज्यों की संख्या क्या है ?

चेन्नई किस समुद्र के तट पर स्थित है ?

अंडमान अवं निकोबार की राजधानी क्या है ?

भारत का सबसे छोटा राज्य कौन सा है ?