Research Article

SEARCHING A SOCIAL NETWORKING SITE USING NATURAL LANGUAGE PROCESSING TECHNIQUE

Aamir Junaid Ahmad[1] and Sabina Priyadarshini[2]

[1]Department of Computer Science & Engineering, Maulana Azad College of Engineering & Technology, Patna, India
[2]Department of Information Technology Birla Institute of Technology, Mesra, India

A R T I C L E   I N F O

A B S T R A C T

We use social networking sites to stay connected with others. Social networking sites provide online platform to build social networks or social relations with other people who share similar personal or career interests, activities or backgrounds. We often want to connect withpeoplewith certain characteristics such as "People belonging to a particular school, college or working in a certain organization" butthe present Social Networking sites do not support such feature. We can only connect with a person who we know by name. Here we present a method to represent and process social network profile information so that we can search people on a social network by querying in English Language. In this paper, we present our work in building a system for translating English sentences into SQL queries for a database system.

INTRODUCTION

Social Networking is an easy way for people to have a social connection with other people with similar interests[1]. These sites serve communication purposes among special interest groups[2], but they do not have a proper searching option where we can look for people or groups having certain characteristics.

If we want list of people with the following characteristics

1. Assistant Professors in the Computer Science Department living in Ranchi
2. Professors in Computer Science Department of BIT Mesra
3. System Analysts working at Infosys in Bangaluru city.

The above query in English Language does not give the results as expected.

Corresponding author: **Aamir Junaid Ahmad**
Department of Computer Science & Engineering, Maulana Azad
College of Engineering & Technology, Patna, India

Query 1. Assistant Professors in the Computer Science department Living in Ranchi.

The output is: Sorry we couldn't find any results for this search

2. Assistant Professors in Computer Science Department who go to BIT Mesra

All of these results are not as expected by the user. In this work we want to add a search option in the social networking sites where we can find people based on their characteristics.

The output is: Sorry we couldn't find any results for this search

3. Assistant Professors in the Computer Science department

METHODOLOGY

Natural Language Interfaces to Databases (NLIDBs) [3] are one of the best options for such situations where users look for information from database and formulate nontrivial andinfrequent queries.The translation of a database query expressed in natural languageinvolves the search of relations that link words of the query (nouns, adjectives, etc.) with elements of the database (tables, columns, etc.), which allow to formalize the query in Structured Query Language (SQL). In this work we have used one sample table, Table 1, which contains all those columns which have been mostly used in the query corpus created from potential user.

Following are the few sample queries which can be answered using our system

- Assistant Professor in the department of Computer Science working at BIT Mesra
- All Professors in the department of Electronics working at IIT Patna
- Accountants in Patna
- Students of St. Michaels working in Delhi
- Girls passed out from Notre Dame Academy living in Patna
- Students of Maulana Azad College working in Bengaluru

We use the following list and data structures in our system.

1. List of cities in the world in alphabetical order
2. List of Countries in the world in Alphabetical order
3. List of Designation with synonym of each designation
4. List of Company/Organization

We have created the following algorithm

1. Accept Natural Language (English Language) query from the user.
2. Remove punctuations (if any)
3. Split the query into tokens
   a. Remove prepositions(on, at, it, from total 94)
   b. Remove pronouns (you, our)
   c. Remove conjunctions (and, but, though)
   d. Remove determiners (the, an, this, some)
4. Identify the words in the user's query which are column names in the table, if any, present in the user's query will be used in the where clause/condition in the select statement.
5. Match each word in the list of cities
   If foundwe append one condition city='word' in the select statement. Since there are only 4037 cities with population more than 100000, the sorted list can be searched in $O(n)$ time.[4] The word is removed from the query.
6. Match each word in the list of countries
   If found we append one condition country='word' in the select statement. Since there are only 196 countries the search will be very fast.[5]The word is removed from the query.
7. Match remaining words in the list of Designations
   If found we append one condition designation='word' in the select statement. As this table is two dimensional containing Similar designations in each row, the complexity will be $O(n^2)$. Since the list is small, for small n we can consider it as $O(n)$
   The word is removed from the query.
   Match remaining words in the list of Colleges[6]
8. If found we append one condition college='word' in the select statement. As this table is two dimensional containing Similar designations in each row, the complexity will be $O(n^2)$. Since the list is small, for small n we can consider it as $O(n)$

Generate the final query and fire it on the database to get the required result which will be displayed to the user.

Implementation

The above algorithm is implemented using HTML, PHP and MySql. HTML is used to design an interface where the user can enter his query in English Language as shown in the Figure1. This is a simple form with a Textbox and a submit Button. When the user submits his query PHP code is used to process the query on the above mentioned technique/algorithm. Based on the user query equivalent SQL select statement is prepared and fired on the database table, Table 1. The output is shown in Figure 2.PHP and MySqlis used as Facebook was initially written using these tools [7]. PHP and MySql provide good connectivity and is easily available.

Table 1 UserProfile

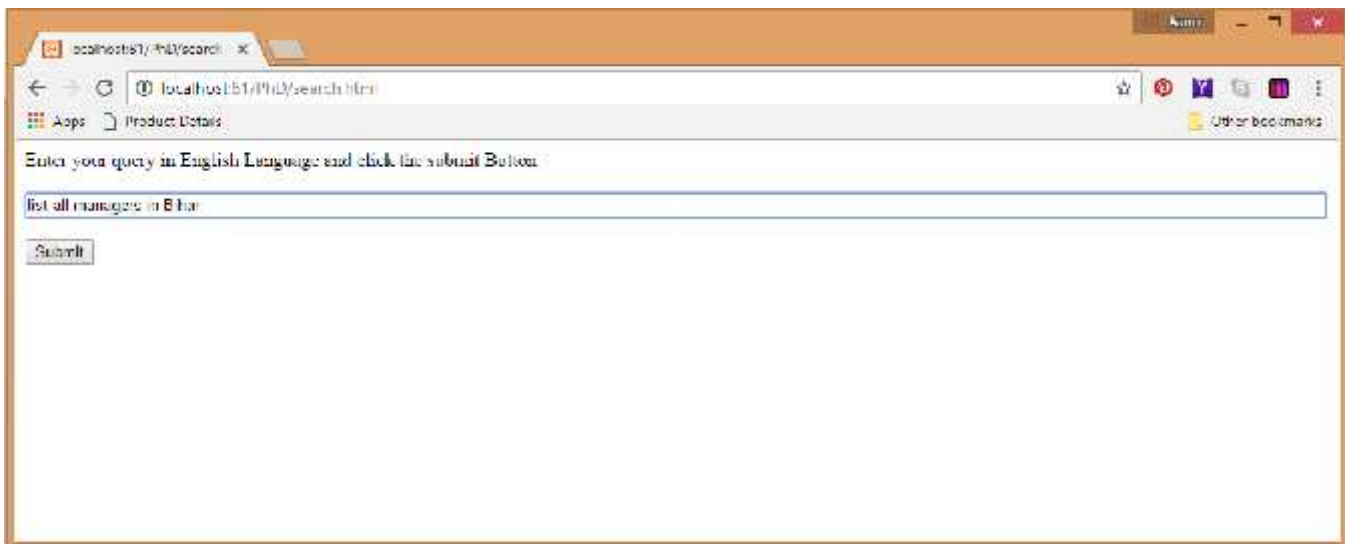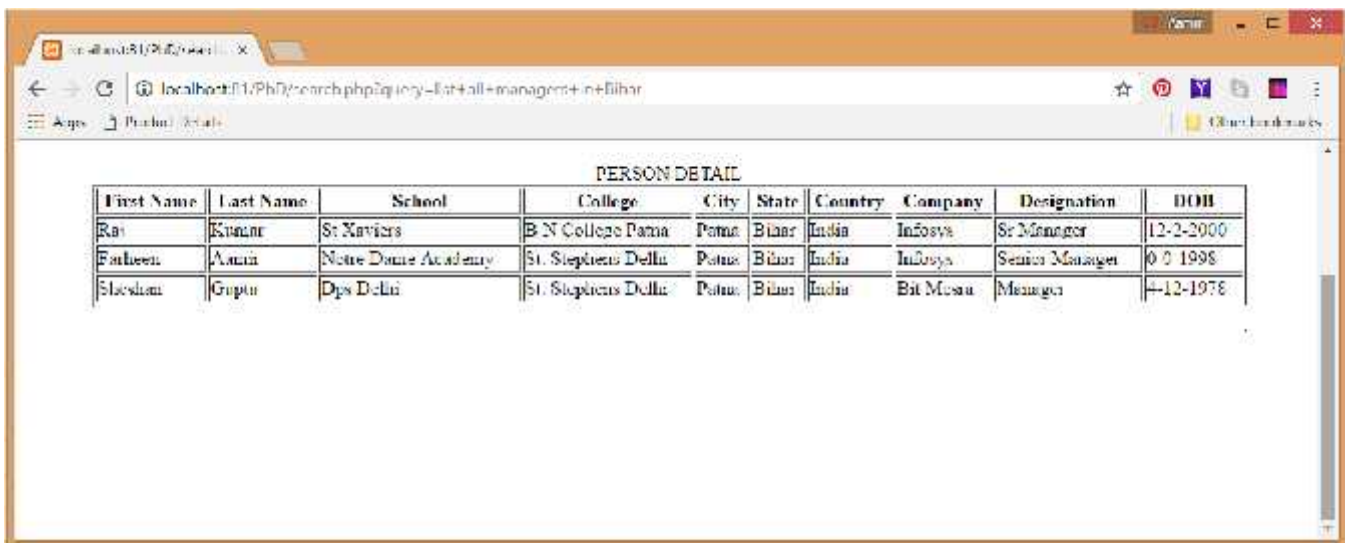| FirstName	LastName	School	College	Day	Month	Year	City	State	Country	Company	Department	Designation
Aamir	Junaid	St. Michaels	BIT Mesra	26	3	1986	Patna	Bihar	India	Maulana Azad College of Engineering & Technology		Professor
Raj	Kumar	St Xaviers	B N College Patna	12	2	2000	Patna	Bihar	India	Infosys		Sr. Manager
Sabina	Priyadershini	St. Josephs	BIT Mesra Ranchi	12	2	2001	Ranchi	Jharkhand	India	BIT Mesra		Asst. Professor
Farheen	Aamir	Notre Dame Academy	St. Stephens Delhi	0	0	1998	Patna	Bihar	India	Infosys	Research	Senior Manager
Farheen	Aamir	Notre Dame Academy	St. Stephens Delhi	20	0	1998	Patna	Bihar	India	Infosys	Research	Senior manager
Farheen	Aamir	Notre Dame Academy	St. Stephens Delhi	20	4	1998	Patna	Bihar	India	Infosys	Research	Senior Manager
Shiva	Malick	St Michaels	IIT Kharagpur	25	7	1997	New Delhi	New Delhi	India	Wipro	Productions	Manager
Sumitra	Giri	St. Josephs Convent Patna	Holy Sumion College	1	4	1996	Pokhra	Nepal	Nepal	Cognizant	Sales	Sales
Dev	Oberoi	AVN Delhi	BIT Mesra	12	5	67	Shanghai	China	China	Maulana Azad College Of Engineering &Technology	Accounts	Accountant
Elroy	Jacoob	Holy Trinity	Patna Womens College	31	7	89	Melbourne	Australia	Australia	BIT Mesra	Computer Science	Programmer
Alastair	Cook	Don Bosco	NIT	21	10	1978	London	England	England			
Ameya	Mathur	Villa Theresa	St. Stephens Delhi	17	6	1986	Mumbai		India	Wipro	Electronics	System Analyst
Leena	Singh	Notre Dame Academy	St. Stephens Delhi	13	12	1990	Paris	France	France	Cognizant	sales	Pilot
Nasreen	Ali	Notre Dame Academy	Oxford University	13	1	1980	amsterdam	Switzerland	Switzerland	Bit Mesra	Research	Manager

Figure 1 Interface Design in HTML

Figure 2 Output

CONCLUSION AND FUTURE WORK

Finally we can say that the time complexity for the above algorithm is of the order O(n). In future we will try to increase the number of columns in the table Table1 to add more attributes which are used in user Profiles.

Acknowledgment

References

1. Aamir Junaid Ahmad and Sabina Priyadarshini. *International Journal of Engineering Research & Technology (IJERT)* Vol. 3 Issue 9, September- 2014

2. Peter Mika, Social Network and the Semantic Web, http://www.springer.com/in/book/9780387710006

3. M. Jose A. Zarate1, R. Rodolfo A. Pazos1, Alexander Gelbukh2, and O Improving the Customization of Natural Language Interface to Databases Using an Ontology

4. list of citites in the world. http://brilliantmaps.com/ 4037-100000-person-cities/ 01/Jan/2017

5. list of countries in the world. http://www.infoplease. com/ ipa/ A0932875. html 01/Jan/2017

6. list of colleges and universities in the world http://www.4icu.org/reviews/index0001.htm01/Jan/ 2017

7. https://code.facebook.com/

****