International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

Phonetic Search in Facebook

Archana M[1], Nandhini S S[2]

[1, 2] Assistant Professor, Department of CSE, Bannari Amman Institute of Technology, Sathyamangalam

Abstract: ***A novel work Phonetic Search in Facebook is presented, which searches the keyword all over the Facebook database and fetches the result based on the pronunciation of the keyword. This system classifies the information retrieved so that the user is able to view the results based on its domain and also provides filter setting during the search and fetches the variants for the keyword and displays the precision percentage for total number of exact and variant match of the keyword. The existing system is Facebook search where it provides search results only for the exact keyword match and also it is not possible for the user to select and find the exact one from the list that is displayed where the user is not supposed to download the result. The proposed system is to access the Facebook database and to group all the retrieved data into the fields such as time, link, location and name. The search to be made is based on the pronunciation of the keyword. It also provides cache memory such that the user will be able to see the results of the search history when the user is offline. The user will be allowed to set whether the search should be made within the pages or public posts or even both. The user requires the Facebook developer API key and the Facebook user account to access the Facebook database to make use of the system. The system will provide the search results along with the variant keyword. The results can also be downloaded.***

Keywords: Phonetic search, Facebook, String Matching, Knowledge Discovery Databases (KDD)

1. Introduction

"Phonetic Search in Facebook" deals with phonetic search in text form, which is a type of string matching algorithm. It describes a proposed solution for phonetic searching of surnames. This solution was designed to improve a search precision and recall for persons with the surnames originated in the standard language. Surnames as the main person identifier, play a key role in present information systems in almost every field. There are various forms of the surname search realization. Searching, which allows finding only exact results in many cases are insufficient and cumbersome. It is also important to find methods to eliminate inexactly defined search queries as well as language barriers associated with them.

1.1 Phonetic Search

Phonetic search is used in applications such as name retrieval, where the spelling of a name is used to identify other strings that are likely to be of similar pronunciation. It also deals about the parallels between information retrieval and phonetic matching [7]. Its demonstration leads to substantial improvements in effectiveness. Phonetic search is used to identify strings that may be of similar pronunciation, regardless of their actual spelling. It guesses the spelling (or is provided with a spelling, which may be incorrect) and uses the guess to query a database of names. The phonetic matching system must then find in the database those strings that are most likely to be of the same or similar pronunciation to that of the query. Algorithms using different approaches for solving these problems exist, but no algorithm can completely solve all types of errors or inaccuracies. Individual algorithms primarily differ in the target language group to which they are intended or typographical errors which are they able to recognize. The phonetic matching approach can reduce input query errors based on a phonetic pronunciation of a concrete language [5]. The phonetic search tries to match strings that have a similar pronunciation, no matter of an actual spelling. The phonetic matching system should then find misspelled name in the database.

1.2 String Matching

Information matching can be divided into several approaches. The first approach is the exact match, when records in the search results must exactly match the input query. The second one is partial or pattern match, when a query contains an uncertainty, which may be masked with wildcard characters or a query only models expected records to match e.g. match using regular expressions. The third approach is called plain string similarity, which is based on comparison of similar groups of letters in two words not regarding to their position. Another approach is trying to eliminate typing errors. It calculates similarity as minimum count of transformations required to transform the first word to the second one. There are also grammatical algorithms, including stemming and synonym matching. And the last one approach is phonetic similarity, represented by language dependent string matching algorithms [2], [3], [10]. The many orders of magnitude faster algorithm opens up new application fields for approximate string matching and a scaling sufficient in real-time. Our algorithm enables fast exact string and pattern matching with long strings or feature vectors, huge alphabets in very large data bases like Facebook with many concurrent processes and real time requirements.

1.3 Various Phonetic Approaches

In most of phonetic name matching algorithms a search request is reduced to a canonic form which is then matched against a database of surnames also reduced to their canonic equivalents. All surnames having the same canonic form as the query surname will be retrieved. The phonetic algorithms result should be appropriately shown to a user like in standard search engines. Hit list should be sorted from "the best" to "the worst" hits according to relevance. According to the way the data are stored in the repository; two techniques can be identified, namely on-line and off-line searching. In off-line search approach a concrete algorithm have stored preprocessed data in repository. In case of phonetic search, algorithms have stored encoded canonic forms of words.

2. Existing Methodology

The existing system is default Facebook search where it provides only search result for the exact keyword match and also it is not possible for the user to select and find the exact one from the list that is displayed where the user is not supposed to download the result [1]. Text mining is known as text data mining or knowledge discovery from textual databases or as an extension of data mining or knowledge discovery from structured databases. In general, it refers to the process of extracting useful information by identifying and exploration of interesting and non-trivial patterns from unstructured text documents. Reference mentioned that text mining uses techniques from information retrieval, information extraction and natural language processing. It also connects with the algorithms and methods of Knowledge Discovery from Databases (KDD), data mining, machine learning techniques and statistics. This collected information is used to gain more knowledge and based on the findings and analysis of the information make predictions as to what would be the best choice and the right approach to move toward on a particular issue [11]. Text mining is the study and practice of extracting information from text using the principles of computational linguistics. It is the process or practice of examining large collections of written resources in order to generate new information. Certainly, AWK and other pattern matching tools can extract information from text files, but these do not fall within the realm of text mining tools. For our purposes, the key areas of text mining include feature extraction, thematic indexing, clustering and summarization. These four techniques are essential because they solve two key problems with using text in business intelligence: they make textual information accessible, and they reduce the volume of text that must be read by end users before information is found.

3. Proposed Methodology

The proposed system "PHONETIC SEARCH IN FACEBOOK" searches the keyword all over the Facebook database and fetches the result based on the pronunciation of the keyword. This system classifies the information retrieved so that the user is able to view the results based on its domain. This system also provides filter setting during the search.

3.1.1 Features of the system

The features of the proposed system are:

- Filter can be set in order to control the count of post to be retrieved.
- User can enter any name of a Person, Brand, Organization to categorize and display results all at one screen.
- User will be able to filter the results from the following sources:
- Public User Posts.
- Pages and Groups.
- Posts within Pages and Groups.
- Results that are generated will be saved in Cache memory.
- The proposed system groups the data based on the category and geographical location,
- This system also provides variants for the keyword that has to be searched.

3.1.2 Advantages of the system

- The system is designed in a user friendly manner.
- It is very efficient to use.
- The system is very reliable to work under stated condition in specific period of time.
- The availability of the system is made user convenient.
- The system provides better results as efficiency is maintained.
- It is platform independent.

System also uses a freeware called JSON (JavaScript Object Notation) which is a lightweight data-interchange format. It is easy for humans to read and write. It is easy for machines to parse and generate. Figure.1 depicts the business logic as PHP since it supports all web browsers and JSON as model. It uses pipe filter architecture. The architecture clearly explains how the system's work flow is maintained.

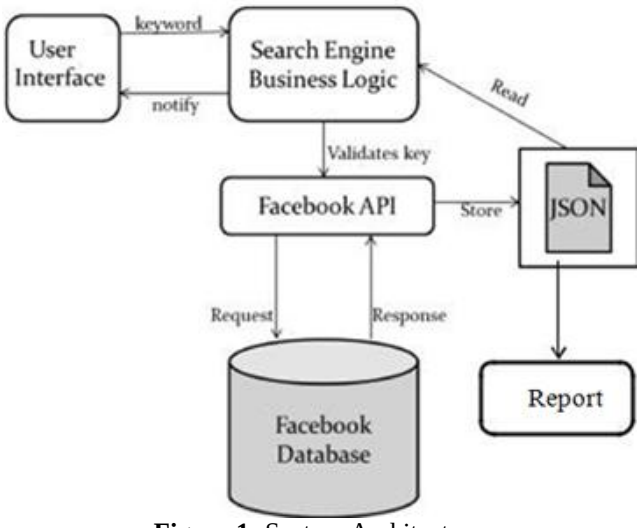

Figure 1: System Architecture

4. System Testing and Implementation

The Testing carried out here are Unit testing, Integration testing and System testing. Unit testing tests all the functionalities individually. Integration testing is done for the whole project working properly. System testing is also done based on the platform dependencies. The following table consists of Test case number, Description, Expected result, Observed Result and the status of the Test case and it describes about test cases and test reports. The system is tested with all possible conditions and they are displayed in table 1.

Table 1: Test cases executed and their corresponding results

Test case	Description	Expected Result	Observed Result	Status
TC-1	Internet connectivity is enabled	Should extract the result after the search	Display all the information in the result page	PASS
TC-2	Internet connectivity is disabled	Should not fetch the results	Check your internet connection	PASS

TC-3	Keyword condition	Keyword should contain only alphabets	Fetch all the results related to the keyword	PASS
TC-4	Keyword condition	Keyword should not contain Special characters	No proper results will be fetched	PASS
TC-5	Count set to 100 posts	Should not fetch more than 100 posts	Fetches exactly 100 posts	PASS
TC-6	Filter set ON for pages only	Should fetch from only pages	Successful results	PASS
TC-7	Filter set ON for pages only	Should fetch information only from posts	Results fetched from posts only	PASS
TC-8	Search history	Should display searched words	Displays search history successfully	PASS
TC-9	Keyword exists	Should fetch results	Display all the reviews related to the keyword	PASS
TC-10	Keyword does not exists	Should not fetch results	Keyword not found	PASS

5. Experimental Results

The Phonetic search in Facebook performs searches over Facebook public posts, and provides tools for analyzing the results retrieved. Search can be customized as needed. Results from Facebook are filtered for interesting information and combined intelligently. The program then offers several tools for analysis and search performance measurement. Multiple searches can be conducted in one session to perform comparisons. This system also suggests alternatives to the search terms, allowing the user to cover all versions of a query name and also query passed is based on pronunciation of it as part of one query. It provides a tool for calculating some essential performance measurements for each search. This can help scientifically validates the findings, and provide an immediate feedback on the quality and reliability of search results retrieved for any search query. The following figures shows the results gathered during the execution of the application.

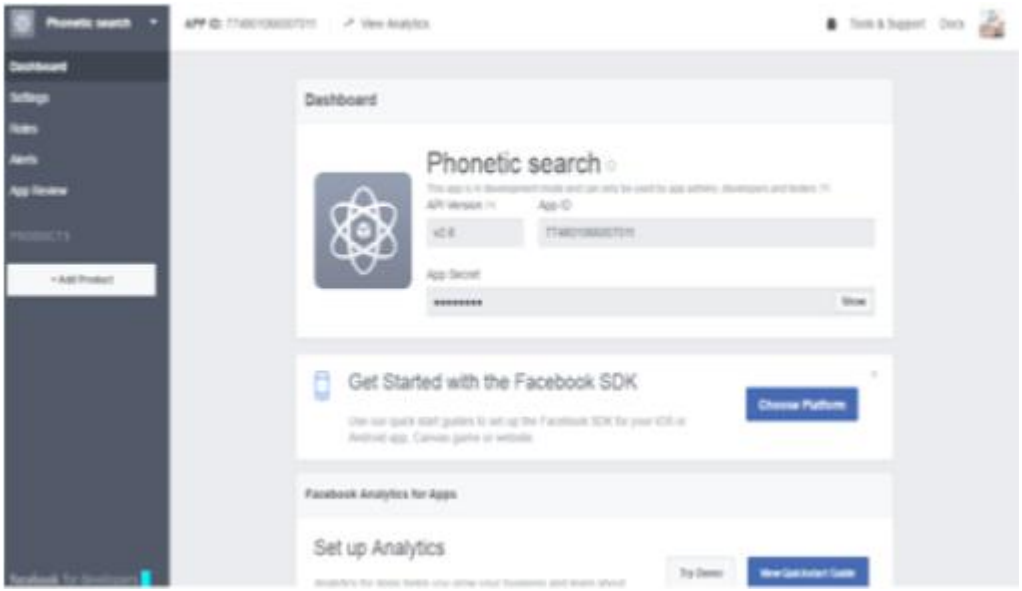

Figure 2: Developers Account for Facebook users

Figure 3: Local host

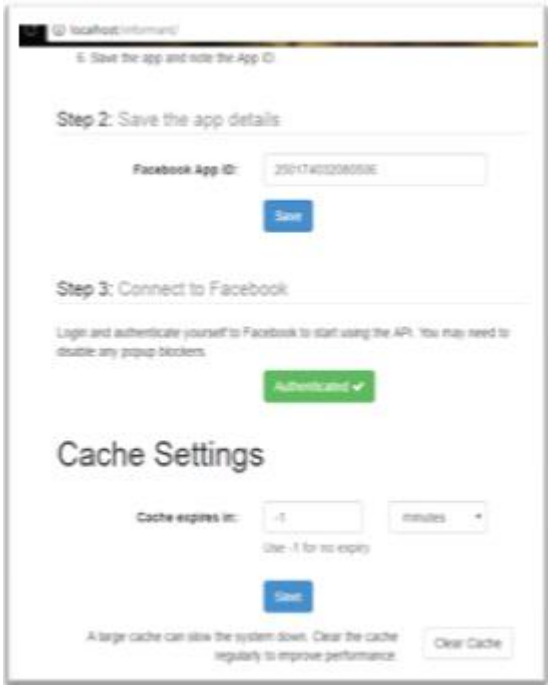

Figure 4: Authentication of App Id

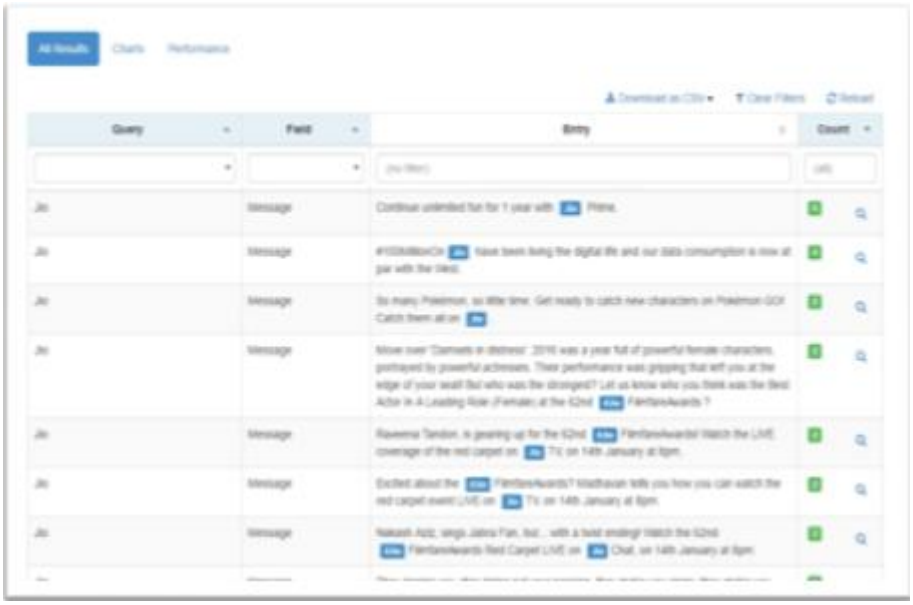

Figure 5: Results Generated for Queries

Figure 6: Performance related to Queries

6. Conclusion and Future Work

This paper proposes a work involving Phonetic search in Facebook that performs searches over Facebook public posts and provides tools for analyzing the results retrieved. Search can be customized as needed. Results from Facebook are filtered for interesting information and combined intelligently. The program then offers several tools for analysis and search performance measurement. Multiple searches can be conducted in one session to perform comparisons. This system also suggests alternatives to the search terms, allowing the user to cover all versions of a query name and also query passed is based on pronunciation of it as part of one query.

References

[1] DoaaMohey& Din Mohamed Hussein, 2016, "A survey on sentiment analysis challenges", Journal of King Saud University – Engineering Sciences,pp.1–12.
[2] Khattab O. Khorsheed , Magda M. Madbouly , Shawkat K. Guirguis, 2015, "Search engine optimization using data mining approach" ,International Journal of Computer Engineering and Applications, Volume IX, Issue VI.
[3] Gayana Fernando & Md Gapar Md Johar,2015, "Framework for Social Network Data Mining", International Journal of Computer Applications, Volume 116 – No. 18.
[4] Sheela Gole& Bharat Tidke,2015,"A survey of Big Data in social media using data mining techniques", International Conference on Advanced Computing and Communication Systems.
[5] Z. Wang, V. J. C. Tong ,David Chan, 2014,"Issues of Social Data Analytics with a New Method for Sentiment Analysis of Social Media Data", Singapore Management University 2014.
[6] Safa Ben Hamouda, JalelAkaichi, 2014, "Social Networks-Text Mining for Sentiment Classification: The case of Facebook statuses", International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 5.
[7] G Nandi & A Das, 2013,"A Survey on Using Data Mining Techniques for Online Social Network Analysis", IJCSI International Journal of Computer Science Issues, Vol. 10, Issue6,No-2.
[8] Dusan Zahoransky & Ivan Pola, 2013,"Rule Based Phonetic Search Approaches", International Journal of Application or Innovation in Engineering & Management (IJAIEM),Volume 2, Issue 5.
[9] Vijay B. Rau & D.D. Londhe, 2012 ,"Survey on Opinion Mining and Summarization of User Reviews on Web", Data Mining Knowledge Discovery,pp.478-514.
[10] Jasha Droppo& Alex Acero, 2010,"Context dependent phonetic string edit distance for automatic speech recognition", Speech Technology Group, Microsoft Research, Redmond, Washington, USA.
[11] N. Azmina M zamani, Siti Z. abidin, Nasirohomar, M.abiden, 2009,"Sentiment Analysis: Determining People's Emotions in Facebook",UniversitiTeknologi 40450 Shah Alam, Selengor Malaysia.
[12] Julie Kane Ahkter& Steven Soria,2009, "Sentiment Analysis: Facebook Status Messages",Stanford University.
[13] JalelAkaichi,ZeinebDhouioui, José López-Huertas Pérez,2000,"Text Mining Facebook Status Updates for Sentiment Classification",Cambridge University Press, New York.