

# Accessing Yellow Pages Directory Intelligently on the Mobile Phone Using SMS

Sunil Kopparapu, Akhilesh Srivastava, Sumitra Das, Rimzim Sinha,  
Meenal Orkey, Vineet Gupta, Jiten Maheswary, PVS Rao

Cognitive Systems Research Laboratory, Tata Infotech Limited,  
Plot No 14, Sector 24, Vashi-Turbe, Navi Mumbai 400 705, India.  
Phone: +91-22-56163251, Facsimile: +91-22-27839926  
<http://www.tatainfotech.com>

**Abstract**— Yellow Pages are directories that source information about various commercial organizations like their addresses, phone contact and other details. These are very useful and are used by individual and other business houses. Until recently, the only way to access these yellow pages directory information was to physically look into a huge hard-copy directory, which was not only laborious but also time consuming and required the user to be familiar with the organization of the directory. More recently, there have been IVR based contact centers that have been set up which can be used by the users to query information. While it is easier than browsing through the physical directory, it still has several pitfalls. The time spent on trying to get the information is quite large and at the end of enquiry one is not sure if one will get the information that one is looking for. In this paper, we propose a novel method (which has been implemented for a major telecom operator) of accessing the yellow pages directory information on the mobile phone by sending a short message service (SMS). The central idea of the proposed method is to avoid any constraint on the way the user can query the yellow pages directory except that it be in natural English. The system, which uses natural language processing (NLP) techniques, understands the intent of the query and intelligently searches the yellow pages directory to retrieve information. This retrieved information is then sent back to the user in the form of a SMS.

## I. BACKGROUND

Mobile phones have made significant inroad into the society in the last couple of years. As we write, there is a large population that is going mobile. While the competition is on for the mobile service provider, one of the ways to retain current subscribers and attract new subscribers is to provide them with value added services (VAS) and at the same time increase the average revenue per user (ARPU). The mobile service provider can retain and increase the ARPU if they provide their subscribers application that are not only innovative but also useful in day to day life[1]. In most of the developing countries, there is a trend to use SMS more than the voice because of economic sense.

Yellow pages directory is a very useful information resource that houses information about commercial organizations. It is very common for a directory to be available for every town or city and it is very often used to get information about the companies. Until recently, the physical yellow pages directory was the only source of information. To get information a user browsed through the directory and got to the information that

he was looking for through the index. There is always the problem of how easy or difficult to reach the information is depending on the organization of the yellow pages directory. Unless very familiar, a user would take effort and time to get to the information. In the recent past yellow pages directories have become accessible through interactive voice response (IVR) systems [2][3]. The user is inconvenienced in the sense that he just needs to make a phone call and request for information; a live agent would search the yellow pages database directory using a series of SQL queries and convey the information back on the phone to the user. While it is a simpler solution, it is

- (a) Time consuming (very often one has to be in the queue listening to advertisement or a very irritation "Please be on hold, your call is important to us and we will get back to you as soon as one of our agents is free to take your call"),
- (b) Expensive for the user (telephone bill for the whole duration of the call including the time taken to get to the live agent!)
- (c) Expensive for the service provider (need to set up a call center and engage  $24 \times 7$  live agents) and
- (d) Highly dependent on the searching skill of the live agent.

In addition acoustic confusion prevail which prolong the interactive session between the user and the live agent because of

- (i) the noisy and low bandwidth telephonic conversation and
- (ii) pronunciation and accent.

In this paper we describe an intelligent and effective system, on the mobile network using SMS that enables searching the yellow pages directory using natural English. Such a system should work on the SMS channel rather than voice because of the cost economics. The system proposed in this paper has the following salient points:

- (a) System should be easy to use
- (b) Should not require the user to remember any specific code or mnemonic to query the directory
- (c) In the absence of exact information (in the database) in response to the query, the system should provide the

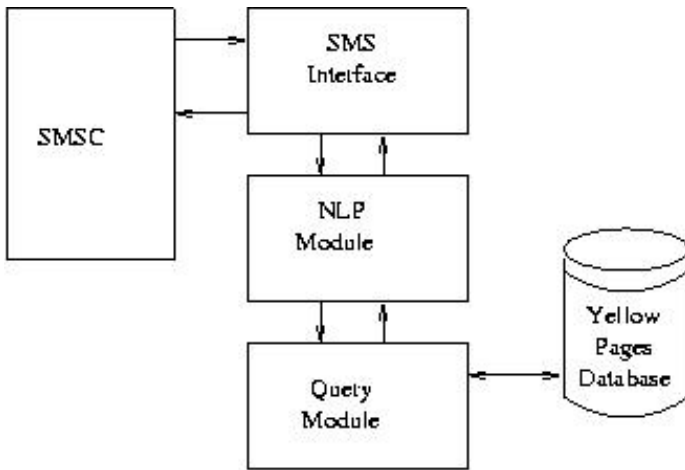


Fig. 1. Overview of the system to query yellow pages on the mobile network.

- user with next close answer in some sense  
 (d) Should cater to SMS lingo and typographic errors that might occur when generating an SMS.

## II. SYSTEM OVERVIEW

The system consists of essentially three main modules (see Figure 1). The first module interfaces with the SMS gateway of the telecom operator and passes the SMS query to the second module, which is the heart of the system. This module which is based on the natural language processing techniques analyses the query in natural English and generates a set of database queries and passes it on to the next module which is the database query module. The output of the query module is passed back to the interface module, which sends back the retrieved yellow pages information back to the user as an SMS. Figure 1 gives the overview of the system that is able to intelligently access the yellow pages directory on the mobile network using SMS.

The SMS module interfaces with the SMSC (SMS center) of the mobile service provider. Its main functionality is to obtain the SMS query from the SMSC and pass on the query to the NLP module, which is the heart of the system. The NLP module understands the intent of the query rather than looking at the query for predetermined set of keywords and generates a list of possible search criterion which can be used to generate an SQL query to extract yellow pages listings from the database. The first search criteria would essentially be what a plain search engine would do namely to use the words in the query itself to search the database, but the subsequent search criteria depends on the actual query itself and can be considered as dilating the constraints placed on the search. The dilation happens in a manner so that not all dimension of the query are diluted simultaneously. We will discuss this in more detail in a later section of this paper.

The system is designed such that it can be easily configured to interface with a SMSC. The SMS interface gets the SMS query and sends it as it is to the NLP module.

The NLP module initially tags each of the words in the query to belong to either the name of the company or the name of a place or a search word (using its knowledge base). In the event of a word not being tagged because of lack of knowledge; it is tagged as unknown and checked for possible spelling mistake using the spell checker module and then tagged appropriately.

The tagged information ( $T_1$ ) is sent to the query module, which forms a SQL query to retrieve information from the database. In the event of no records being returned from the database, the NLP module generates another tagged list ( $T_2$ ) which is used by the query module to search the database. This process is continued until one or more records are returned from the database to the query posed by the user. The generation of the tagged list ( $T_1, T_2, \dots$ ) depends on the initial query and is such that the SQL query has a higher probability of attracting records in the database.

A prototype of the system that can access the yellow pages directory for the Mumbai yellow pages directory was deployed for a major telecom operator in India. The next section describes the key advantage of using the NLP based strategy that allows the user the freedom of how and what they ask. The system answers intelligently (as a normal human would do) by giving answers when present in the database else giving approximate but close answers in the absence of an exact answer being available in the database.

## III. NLP ADVANTAGE

Designing the strategy of generating  $T_1, T_2, \dots$  from the initial query of words is crucial and helps in digging out information from the database in the event of information being absent for the initial query words. This strategy gives our system an edge over using an ordinary keyword based search strategy. The strategy information is part of the NLP module (Figure 1). In addition the NLP module is supported by a taxonomy tree<sup>1</sup>.

We establish the advantage of using such a strategy through actual examples<sup>2</sup>.

**Example: 1:** Suppose a user is looking for a **Studio in Eastern Andheri**<sup>3</sup>. An ordinary search strategy would fetch: [1] EASTERN TRADERS, ANDH [2] EASTERN ELECTRONICS [3] M K EST, 73, M K EST, ANDH (E), 28590034, records from the yellow pages database while the NLP based strategy would enable extract appropriate and exact information, namely, [1] M K EST, 73, M K EST, ANDH (E), 28590034 [2] GEMINI STUDIOS, C/3, M.I.D.C., ST NO 11, ANDH (E), 28229933 [3] KAMAL AMROHI STUDIO, JOGESHWARI VIKHROLI LINK, ANDH (E), 28208026 [4] CHANDIVALI OUTDOOR STUDIO, CHANDIVALI RD, ANDH (E), 28521097

<sup>1</sup>A tree structure, that captures relationship between different words. Traversing through the tree one is able to relate different words. In literature it also goes by the name ontology tree.

<sup>2</sup>For Mumbai yellow pages directory

<sup>3</sup>Andheri is a suburb in Mumbai

While an ordinary search strategy would produce any results with any of the words in the query (studio, eastern Andheri) as the keywords, the NLP based strategy would be able to return better search results by understanding that the query intend to search for a studio in 'Eastern Andheri'.

**Example: 2:** For a query **Want to have Meduvada in Juhu**, an keyword based search engine would not return any results while the NLP based system would return [1] UDIPI SHREE KRISHNA, JUHU CHURCH RD, JUHU, 26713178. Following may also be useful: [2] THE SEAFARER REST, LIONS, CNTRL JUHU BEACH, JUHU, 26162839 [3] SAPPHIRE, THE EMERALD, JUHU TARA RD, JUHU, 26611150 [4] SUBURBIA REST & BAR, GAYLAND HTL, JUHU TARA RD, JUHU, 26170999

**Example: 3:** Suppose a user queries **looking for DTDC Couriers in Sanpada**. In the absence of a DTDC courier in Sanpada the ordinary search strategy would produce no output, while a NLP based system would give close and appropriate records extracted from the yellow pages directory [1] DTDC STALLION ENTPS, STALLION ENTPS, A 31, VASHI PLAZA., SEC 17., VASHI<sup>4</sup>, 27894652 [2] DESK TO DESK COURIERS, 7, GR FLR, AMBASSY CENT, NARIMAN POINT, 56311357 [3] DESK TO DESK COURIERS, GALA NO 15, 1ST FLR, MEHTA STATE, NDHERI KURLA RD, ANDH(E), 56943478 [4] DTDC COURIERS, 15, MEHTA EST, AND-KURLA RD, ANDH(E), 56943477

**Example: 4:** For a query, **Citibank ATM in Vashi**, the keyword based search strategy returns no results (because of a Citibank ATM being absent in Vashi). On the other hand the NLP strategy based system is able to give information that is useful while suggesting that that there is no perfect fit for the query posed. It says No perfect fit for CITIBANK ATM IN VASHI. Hope this helps: [1] CITIBNK, PANCHEEL ARCD, SEC 5, AIROLI<sup>5</sup>, [2] INDUSIND BNK LTD, MANEK CPLX, SEC 29, VASHI, [3] UTI BNK SHP 1, PL 17, SHIV DARSHAN, SEC 4, VASHI [4] UTI BNK WARDHAMAN CHMBS, PL 84, SEC 17, VASHI, 27660066

**Example: 5:** For the query **Breakfast in Taj** while keyword based search would results in no record or any record having Taj as the company name, NLP based strategy results in correct results because of the ability to relate breakfast to a place which serves food. The records returned by NLP based strategy are [1] TAJ GROUP OF HTL, MANDLIK HSE, MANDLIK RD, COLABA, 22022626 [2] TAJ GROUP OF HTLS, MANDLIK RD, APOLLO BUNDER, COLABA, 56653366 [3] TAJ PRESIDENT, 90, G D SOMANI RD, CUFFE PARADE, 56650808 [4] TAJ LANDS END\* REGENT, LANDS END, BANDSTN, BDRA(W), 5668123

**Example: 6:** For a query **Buying Jeans in Andheri** ordinary search strategy produces no results while the NLP based strategy gives [1] IMAGE APPARELS P LTD, ARVIND CHMBS, WERN EXPRESS HIGHWAY, ANDH(E), 28224892 [2] LIVE IN JEANS, C-6, MIDC, RD NO 22,

ANDH(E), 28252127 [3] APEX JEANS WEAR, 9/F, NAND-JYOT INDL PREMISES, ANDH KURLA RD, ANDH(E), 28511891 [4] SINGAPORE OLLECTION, DN RD, ANDH(W), 26209109

**Example: 7:** For the query **Cable operator in Vashi**. While the ordinary search strategy produces no results, the NLP based strategy gives [1] SSV CABLE P LTD, 9, NR INDIAN BNK, LANDMARK CHS, SEC 14, VASHI, 27664073 [2] AASHISH CABLE NET INDIA P LTD, SEC 9 A, GURAV HALL, VASHI, (O)27655535. In addition it also list [3] SEVEN STAR CABLE, SHP-3, MINI JEWEL, OPP GTB BNK, SEVEN B'LOWS, ANDH(W), 26362675 and [4] UCN HATHWAY CABLE, 2ND FLR, STRAND CNMA BLDG, COLABA, 22812994 suggesting them as a possible alternative answers to the query.

These examples clearly demonstrate the value add of using NLP based strategy to retrieve information from the database. In all the cases, demonstrated, the ordinary search strategy fails because of either the absence of the information in the yellow pages directory or the inability to extract more information from the query instead of treating all the query words as being key search words.

#### IV. COMPARING WITH OTHER DIRECTORY SEARCH PRODUCTS

To the best of our knowledge there does not exist an equivalent product in the market. The closest one we could find is a service is provided by TM Cellular Sdn Bhd [4] (TMTouch) and allows query for business (company name and classification), government, police department and embassy only for the Malaysian yellow pages directory. The Malaysian yellow pages directory provides for its subscribers to query the yellow pages directory on their mobile phone using SMS. There are constraints (have to use one of the predetermined keyword like hotel, bank etc) on the way the query is formulated and the information that can be searched.

Several other operators' worldwide provide what is called the city guide as a value-added service. It is usually a WAP enabled service that allows one to browse information statically stored as a set of hyper-links. So to get some information one will need to browse the information available on that page and then click on the information that he is seeking. After several such browses and hyper-link clicks he is presented the information (provided that the information is available) that he is looking for! In contrast, to a typical city guide, our application, allows a person to send the query as an SMS to intelligently query Yellow Pages Directory. Information in response to a query is searched and retrieved by the system (without the user being required to manually search and click several hyper-links) and the exact information that the user is searching for is sent back to him in the form of an SMS. Economically speaking<sup>6</sup>, browsing through the city guide application, is an expensive proposition because

<sup>4</sup>Note that Sanpada and Vashi are adjacent suburbs in Mumbai

<sup>5</sup>Note that Airoli is one of the adjacent suburb of Vashi

<sup>6</sup>In the Indian context

- Sending and receiving an SMS is less expensive compared to WAP browsing with most telecom operators
- Lower end phones[5] are capable of sending and receiving SMS messages while for WAP browsing one will need a more expensive phone and
- Time spent for browsing and clicking is more than the time taken to send the SMS
- In the absence of exact information, in the database, a city guide would provide with no results while our application provides next best results.

## V. RESULTS

The system was put to test by selected users, none of whom were associated with the development of the system. The queries posed by them were classified as good questions and bad questions<sup>7</sup> for the purpose of analysis. The bad questions typically were out of domain of yellow pages or too general or absurd. Only the good questions were analyzed.

Question Types	%
Good Questions	87.7
Bad Questions	12.3

TABLE I  
DISTRIBUTION OF QUERIES

The users were themselves asked to rate the response into one of the categories (satisfactory response by the system, not satisfactory and can not say). The performance of the system is shown in Table II.

Answer Types	%
Good Answers	86.0
Incorrect answers	12.7
Can't Say	1.3

TABLE II  
ANALYSIS OF THE RESPONSE

## VI. CONCLUSIONS

In this paper we have developed an intelligent system that is capable of accessing information from the yellow pages directory. The system is user friendly because it allows the user to access information in the yellow pages directory by posing the query in natural English. The cost of accessing information is low because it is via an SMS, which is the cheapest mode of communication on the mobile network in developing countries. One of the salient features of the system is its ability to give the next best answer in the absence of exact information not being available in the database. A working prototype was developed and deployed for one of the major telecom operator in India.

## REFERENCES

- [1] PRWeb, "Happy VAS customers translate directly into increased ARPU and profit," in <http://www.prweb.com/releases/2003/2/prweb58656.htm>, 2003.
- [2] YellowLine, "Infomedia Yellow Pages," in <http://www.yellowpages.co.in/>, 2004.
- [3] JustDial, "Talking Yellow Pages," in <http://www.justdial.com/>, 2004.
- [4] TMTouch, in <http://www.tmtouch.com.my>.
- [5] C. S. Labs, "SMS (short message service) -technical overview," in <http://www.cswl.com/whiteppr/tech/sms.html>, 2004.

<sup>7</sup>Example bad queries - (a) Out of the yellow pages domain e.g. Cricket score (b) Absurd queries e.g. AAAAA (c) Too General Queries e.g. India