Research Journal of Computer and Information Technology Sciences _____ISSN 2320 – 6527
Vol. **5(5),** 12-14, July **(2017)** *Res. J. Computer and IT Sci.*

Short Review Paper

A study on text auto completion for easing search of information

Isha Verma* **and Monika Verma**
Bhilai Institute of Technology, Durg Department of Computer Science and Engineering, Durg, CG, India
iverma13@gmail.com

Available online at: www.isca.in
Received 10th April 2017, revised 18th July 2017, accepted 20th July 2017

Abstract

Text prediction has multi-dimensional applications, ranging from easier database access to resolving of cognitive disabilities. The tools for such applications exist, but are not tolerant enough of error to cater the needs of masses. Moreover, these systems depend upon long duration initial queries which are used to build word lists that aid in query expansion. Such systems do increase their vocabulary of generic words used by users but cannot aid in formulation of words that are rarely used. In this paper, various factors affecting auto-complete, the aid provided by it to the dyslexic patients, its advantages and application have been discussed. In every domain where lexicography has to dealt with as an extension of information retrievals, auto-complete is a necessary tool for assisting people with query generation, so that any inherent cognitive disability or lack of proper vocabulary does not affect the retrieval of information.

Keywords: Auto-complete, Communication rate, Text prediction, Keystrokes, Dyslexia.

Introduction

The term text prediction means a set of computer program and algorithms, which help users, edit text with higher efficiency. Their integration can be observed in a broad spectrum of applicative programs such as email editors, texting and web browsers, found in computers and mobile devices. They can be classified into two types: auto-completion and auto-correction[1].

Auto-completion, refers to the prediction and completion of a word, or a phrase, as soon as we start typing. The methodology is based upon the selection of the most likely word in a confined set of probable completions. The text prediction task consists of editing text with the minimum number of keystroke feasible. Possibly, the best known illustration is the T9 patented technology, which is widely employed in the initial generation of mobile phones[1].

In auto-correction (or text replacement) the text which is already typed is completed or replaced. It is used to either correct text containing grammatical errors or to select the apt word in a dictionary. Although in this case the process starts "after" finishing the word insertion, it can be considered a type of text prediction. As a matter of fact, the word to be replaced is generally typed mistakenly, and the (predicted) replacement word is what the user wants to enter. Mostly, text prediction is reduced to a word selection task[1].

In the upcoming sections, the factors affecting auto-complete, advantages, applications and future scope have been discussed.

In every domain where lexicography has to be dealt with as an extension of information retrievals, auto-complete is a necessary tool for assisting people with query generation, so that any inherent cognitive disability or lack of proper vocabulary does not affect the retrieval of information.

Factors affecting auto-complete

A set of usability tests were conducted on the use of auto-completion in various academic research scenarios. From the study the following points contribute as factors for its implementation[2]. First, Spelling: Auto-complete is helpful when a person has idea about the word but doesn't know the correct way to spell it. Second, Number of key strokes: It decreases the number of keystrokes while typing. Third, focusing known items: It assists when we want to find a specific item and we have a partial citation about it. It also helps in narrowing down the ideas if the user is at a later stage in the research process. One can also consider new similar topics which he might have not thought about. Fourth, Confidence: It gives users the confidence about any novel topic which they are looking for is not imaginary and that details about the same can be found somewhere. Fifth, Speed: It is observed that auto-complete is a time-saver and it does so when a user clicks on the suggestions[2,4].

Effect of auto-completion

The effect of auto-completion has been multi-varied ranging from augmenting communication rate to acting as an aid to users with cognitive disabilities.

A study was conducted to understand the effects of auto-completion on the communication rate. When a text is typed, its structure provides a framework in the form of a context, when

this is seen in the light of a predetermined lexicographical structure, a strongly accurate word list can be generated.

The framework based on the recentness of the words competes with the one word retrieval framework, especially when a potential word list is absent. The next generation word completion creates lexicons using a three word model combined to a back-off framework, and a usage of one word framework is employed for the each initial word of every statement. The aforementioned lexicological framework is obtained based on the telephone transcripts.

If the aforementioned (three word model) framework fails to beget results the system reverts to a framework based on the recent occurrences of the words[4].

The information searching behavior may be affected, when used by the dyslexic patients. If the typical spelling errors made by them can be corrected using a list of spelling mistakes typically made by dyslexics and correct them using real time spell checkers to give the users more autonomy over the query choices. But, since these technologies already exist, the resulting inference tells us that there is a need for a different search interface to decrease the keystrokes and the tedious amount of result lists generated per query fetched because of some traits of dyslexia. For such users, diminished naming skills and spelling errors is problematic hence making the query formulation a big challenge. Naming skills call attention to associating visual and verbal information by assigning appropriate names to, for example, numbers, objects or colors. Moreover, diminished short-term memory and slow reading may pose problems with skimming result lists and document assessment. Thereby, increasing the effort needed to solve a search query[3].

In the study, dyslexic users were taken into consideration. The participants consisted of 40 volunteer under graduation and Post graduation students, 20 were diagnosed with dyslexia and 20 were non-dyslexic formed the control group. The participants were asked to solve 10 predefined tasks with the help of Google[3].

As a remedy for the cognitive shortcomings of the people who suffer from dyslexia, query expansion was hypothesized as a solution assuming that the search duration and submitted number of queries are same in the case of using auto completion, and this was supported by empirical results. It was also found that the compensation for the reading and writing difficulties rendered the dyslexics capable of making information searches with as much ease as the people participating in the control group[3].

An amusing feature of Google is that, it offers supplementary query-building support. For Example, if the search engine presumes that a query is miswritten, Google returns results for what is considered the apt query. Moreover, to draw attention to likely spelling errors, a red line is displayed below words categorized as misspelled while the user enters the query[3].

Advantages

Firstly, the study conducted with a small number of Augmentative and Alternative Communication (AAC) users yield some powerful results i.e. the word prediction technique can increase AAC communication rate[4]. In the work, the independent variable is the method of text entry used: i. one-by-one letter typing with no word prediction, ii. one-by-one letter typing corroborated with word predictions produced by a basic prediction method, iii. one-by-one letter typing supplemented with word predictions which were developed by an advanced prediction method[4].

The advantages of auto-complete are: First, Speed, using auto-complete helps users to fill a form quickly. Second, Communication rate: Auto-complete increases the communication rate thereby, incrementing the speed of communication[4].

Applications

A Mobile Real-Time Query Expansion Designing Phase Builder: The results show that offering Real time query expansion (RTQE) leads to better quality initial queries, more engagement in the search, and an increase in the uptake of query expansion. The quality of queries submitted to Information Retrieval (IR) systems directly affects the quality of search results generated by these systems. For this reason the issue of how to improve search queries has been of great interest in IR research. The major focus was to advance the query expansion by facilitating the selection of indexed words even though a similar or identical query was not indexed. To achieve this, the word register and the framework employed to retrieve words must make predictions quickly and such a function must be served especially when the internet connection is slow. When this aforementioned conditionality was added to our limitations with the fact that upgradability our index with the queries generated by the users was a logical twin step, an on-site memory was to be incorporated along with a framework to pinpoint a particular word.

If keystrokes are emphasized upon, different elements of lexicology involved will not be excessively smoothened to fit around a certain central element, and will therefore help increase the randomization of the sample of the participants involved in the auto-complete experiments.

Also to reduce the keystrokes would be a second focus. Traditional methods of text forecast use a lexicon that is not specifically designed to suit the auto-complete function[5,6].

The major focus was to advance the query expansion by facilitating the selection of indexed words even though a similar

or identical query was not indexed. To achieve this, the word register and the framework employed to retrieve words must make predictions quickly and such a function must be served especially when the internet connection is slow. When this aforementioned conditionality was added to our limitations with the fact that upgradability our index with the queries generated by the users was a logical twin step, an on-site memory was to be incorporated along with a framework to pinpoint a particular word.

If keystrokes are emphasized upon, different elements of lexicology involved will not be excessively smoothed to fit around a certain central element, and will therefore help increase the randomization of the sample of the participants involved in the auto-complete experiments.

Also to reduce the keystrokes would be a second focus. Traditional methods of text forecast use a lexicon that is not specifically designed to suit the auto-complete function[5,6].

Snip Suggest: Context Aware Auto completion for SQL: The Sloan Digital Sky Survey (SDSS) is a famous example of this shift toward data-intensive scientific analytics and SQL[7].

Testing of a hypotheses by evaluation of queries over massive datasets. Snip Suggest's present abilities include suggestion of tables, views, and table-valued functions in the FROM clause, columns in the SELECT clause, predicates in the WHERE clause, columns in the GROUP BY clause, aggregates, and some support for sub-queries.

All queries, accurate or inaccurate, are logged by the Query Logger.

This is a hindrance for Snip Suggest because of the massive workload its response time becomes deteriorated[7].

Future enhancements

People suffering from dyslexia and other mental disorders that hinder the reading and writing abilities, usually suffer from the loss of ability to use accurate vocabulary due to their inability to accurately perceive spellings, and if the already existing auto-complete features could solve this problem it would have already been helpful, but since that is not the case, a different system to mitigate this problem is necessary. An auto-complete system that has a massive spelling error tolerance, this can enable the dyslexics and non-dyslexics alike, to construct more accurate queries.

This feature can be used to make data access in great amount of data holding databases even more accurate for applications in operation of databases designed for quantum computing which will become the next best data mining system due to it exponential computing power.

Conclusion

The study of Auto-completion in it's various forms had shown us several inconspicuous gaps in possibilities and implementation. To realize all the potential of auto completion the aforementioned methodologies must be rigorously pursued, and in the process a tool as intelligent as young human can be envisioned with various fail-safes employed to keep it's efficacy intact.

References

1. Valitutti A., Toivonen H., Gross O. and Toivanen J.M. (2012). Decomposition and Distribution of Humorous Effect in Interactive Systems. In *AAAI Fall Symposium: Artificial Intelligence of Humor,* 96-100.

2. Ward D., Hahn J. and Feist K. (2012). Autocomplete as a research tool: a study on providing search suggestions. *Information Technology and Libraries*, 31(4), 6.

3. Berget G. and Sandnes F.E. (2015). Do autocomplete functions reduce the impact of dyslexia on information-searching behavior? The case of Google. *Journal of the Association for Information Science and Technology*, 67(10), 2320-2328.

4. Trnka K., Yarrington D., McCaw J., McCoy K.F. and Pennington C. (2007). The effects of word prediction on communication rate for AAC. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Companion Volume, Short Papers*. Association for Computational Linguistics, 173-176.

5. Paek T., Lee B. and Thiesson B. (2009). Designing phrase builder: A mobile real-time query expansion interface. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services,* ACM, 7.

6. White R.W. and Marchionini G. (2007). Examining the effectiveness of real-time query expansion. *Information Processing & Management*, 43(3), 685-704.

7. Khoussainova N., Kwon Y., Balazinska M. and Suciu D. (2010). Snip Suggest: Context-aware autocompletion for SQL. *Proceedings of the VLDB Endowment*, 4(1), 22-33.