Journal of Engineering Sciences
& Information Technology
Issue (1), Volume (2)
March 2018
ISSN: 2522-3321

المجلة العربية للعلوم و نشر الأبحاث
Arab Journal of Sciences & Research Publishing
AJSRP

Data Vocalization
Enhancing Voice Output of Relational Data Base

Afaf Khaleel Edinat **Fatima Omar Haimour**
King Abdullah II School for information Technology||University of Jordan || Jordan
Duha 'Moh'd Basam' Qutishat
Prince Al-Hussein bin Abdullah II School for Information Technology || Hashemite University|| Jordan

Abstract : Recent advances in automatic speech recognition and natural language processing have led to a new generation of robust voice-based interfaces. Yet, there is very little work on using voice-based interfaces to query database systems. Current trends towards voice-based interfaces motivate the problem of data vocalization, a complementary problem to data visualization. So, the research on data visualization aims at finding the best way to present data via visual interfaces. In this paper we will discuss method to classify the users and their interest of how to select between the simple voice outputs row by row and concise version to find the good policies to select between simple and concise version is an interesting direction. This work is considered as extend to work proposed by Trummer et al.[1].

Keywords: Data Vocalization, Relational Data Base, Speaking Time Customization

1- Introduction

Querying data in relational database is challenging .SQL is the standard query language for relational data base; SQL is very difficult for users without technical training. As shifting the users of database form expert's users to non-experts, designing user–friendly query interfaces will be very important goal in database community [1]. To achieve this goal there are many researches on data visualization which aim at finding the best way to present the relational database in efficient way to the users via voice interfaces [2].

Recent advances in automatic speech recognition , natural language processing and emerging new tools with devices enabled new generation of voice- based interfaces, such, Apple's Siri ,Google Home, Amazon Echo and voice based SQL interfaces) when used together, the relational database systems, voice- based interfaces providing intuitive way to query and consume data[3], these voice-based interfaces are superior in many situations where a context switch is either impossible or could cause distraction. The benefits of the voice-based systems providing the general accessibility of information and supports the disabilities people that prevented from using screens, keyboard and gestures, and by having the voiced based interface for data base querying. This will allow these people to interact with database

directly without having to use cumbersome and inefficient workarounds such as format ice, eye tracking in order to use SQL or any other query interfaces.

Despite the advantages of using the voice based interfaces to query data base systems, still there are many challenges and barriers of this system .In this paper we will try to address these barriers and challenges which we need to remove, and to increase the expansion of these systems we should to improve it to attract the users, and we will describe one of the most important voice-based interfaces which recently proposed call Query-by-Voice (QbV).

The rest of this paper is organized as follows: The next section presents related work; section 3 describes in details the voice-based interface. Section 4 presents the system architecture, section 5 Query interface. Section 6 represents Speaking Time Customization. Finally section 7 includes the conclusions and future work.

2- RELATED WORK

Prior work on constructing natural language interfaces to database has been studied for several decades [3], and on generating natural language descriptions of data focus on producing written text [4], other papers focus on the voice output which is subject to specific constraints. It has to be extremely concise(as opposed to written text, user cannot skim text to identify relevant part quickly), and has to respect memory limitations of the listener(as opposed to written text , user cannot easily re-read prior passages)[3], prior work present the relational database as data signification as opposed of data vocalization focuses. There are many approaches introduced in prior work about information presentation in spoken dialogue systems are typically specific to scenarios [5] .

There are many papers proposed several methods to improve the communication between users using natural language and voice-based systems, such as system Question-And-Answer classification technique for constructing and managing spoken dialog system (QADB), By classifying the question examples in the database into some clusters using the probabilistic Latin Semantic Analysis (PLSA), an appropriate question example can be found more quickly than when using the conventional method [6].

One of the papers focuses on optimizing voice output of relational data to present data in most efficient way and to make voice output as concise as possible to avoid exceeding user attention span, also to respect the limitations of the user's short term memory, by minimizing the speaking time under constraints that reflect the particularities of voice output, to make the speech easy to remember and simple understand for users [2].

Voice-based interfaces need more attention and improvement to introduce to the users with an ideal way, there are many challenges and barriers. One of the most important challenges is how can we categorize the users and their interest, which users need to use these systems, if there is specific users who attracted to these systems such as, disabilities people, children, businessmen or people who don't have

enough time to search for specific information using their hands (hands-free) or texting because of their busyness.

The speaking time is another important challenge because we need to determine policies to categorize the users by selecting simple voice output row by row, or concise version of voice output, because there are many users who prefer the concise version who don't have enough time or who are looking for exact and particular information in a simple structure, easy to remember and understands.

In this paper we will discuss method to classify the users and their interest of how to select between the simple voice outputs row by row and concise version.

3- Voice –based interface

There is a very little work on using voice-based interface to query database system, a new querying and interaction paradigm they call Query-by-Voice [1], using a proof-of-concept system called Echo Query, the query interface of Echo Query is inspired by casual human-to-human conversation.

The main features of the voice – based interface of Echo Query are:

- **Hands-free Access:** Echo Query does not require the user to press button or start of an application using a gesture. Instead, user can interact with database using their voice at any time [1], this features will help many users to complete their work in case of busyness, and they can do several works at the same time.
- **Dialogue-based Querying:** Echo Query provides a stateful dialogue-based query interface between the user and the database where (1) users can start a conversation with an initial query and refine that query incrementally over time, and (2) Echo Query can seek for clarification if the query is incomplete or has some ambiguities that need to be resolved. The opposite of the traditional database systems which provide (i.e., stateless) a query interface.
- **Personalizable Vocabulary**: provides the ability for users to formulate the queries using their own terms which might be different from the schema elements of database.

4- system architecture

Echo Query is built using a middleware approach over an existing relational database that provides a SQL interface, the database system is used to efficiently store and query the data, as shown in (Figure. 1).

Echo Query implements three different intents: a Query Intent, a Refine Intent, and a Clarify Intent. In the following sections we will discuss briefly these intents.

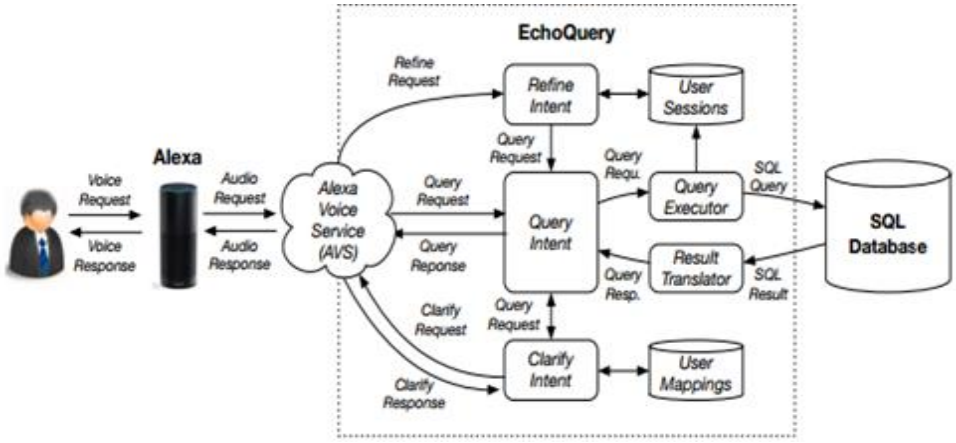

Figure1: Echo Query Architecture[2]

5- Query interface

Echo Query's interaction model is composed of different intents, which implement the conversation between the user and the echo query system[2].

5.1 Query Intent:

This intent is used by Echo Query to start a conversation with the user, the Query Intent adds the missing information that not specified by the user. Afterwards, the completed query request is translated into executable SQL query using the query translator [2].

5.2 Refine Intent:

This intent is used by Echo Query to answer a follow-up query in a conversation between the user and the database. The purposes of this intent is as the following:

- Extract some particular details for result of previous query.
- Modify some parts of the previous query.

The Refine Intent therefore stores the last query of each user session [2].

5.3 Clarify Intent:

If some information to actually execute a query request cannot be automatically inferred or some information in the query request is ambiguous, the Query Intent calls the Clarify Intent. To ask the user for more information to disambiguate the query request by using a voice – based conversation.

A second function of the clarification intent is that learns the user terminology during the course of conversation. Therefore, after asking the user clarification, Echo Query stores the mapping of user terminology to schema elements [2].

6- Speaking Time Customization

A lot of studies show that that it is possible to extract information regarding the personality of individuals from similar datasets or users log files by using even simple data mining techniques[7][8]. Despite its simplicity we hope that it will encourage the development of more sophisticated models to extract Personality Information. We believe that the information extracted can be used to gain more concise data in accordance to the user needs.

In data vocalization researches all aims are to minimize speaking time without taking into consideration the nature of the listener preferences whether the user prefer summarized or detailed taking into considerations some constraints to avoid cognitive load on the listener in contrast with written texts.

We are trying to study data vocalization in different perspective. We are also studying the question of how to optimally improve vocalized output by generating results of a degree of detail that can be optimal to each individual user.

In our proposed model we try to take some of the listener characteristics that can be extracted from his/her log file trying to extract the needed constraints instead of being predefined, in other word customizing the speaking time in accordance to the listener characteristics while observing the conciseness of the results (Figure. 2) shows the our suggested model outline.

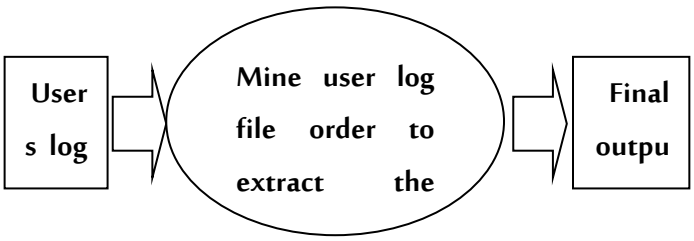

Figure2: proposed Model Workflow

In addition to the user 's log file we can also rely on a number of predefined characteristics that can be used in determining out our intended constraints, but using these characteristic will be limited to queries of known domains in order to be able to pick the characteristics that can be considered a good feature for our final output such as genre information (gender and age), academic level, medical history, fields of interests, or even informal data like hobbies and sport activities. (Figure 3) shows the pseudo code for the proposed model.

Pseudo code of Customized Constraints Extraction

1) Input: relation to vocalize.
2) Search space :Mine user log file
3) Extract customized Constraints
   - Context size
   - Categorical value domain
   - Domain size
   - Numerical value domains

   Upper bound<= lower bound * p

   (Where p is the precision)
4) Objective: customize speaking time according to user constraints.

Figure 3: Pseudo code for the suggested model

According to these constraint instead of being predefined they will be determined according to user log file, we can determine the characteristic of the users from their interests therefore ,give them higher precision, for instance if the user age attribute has value greater than 70 and he inquiring a restaurant our suggested algorithm will give location high precision and assign the context size to low value.

6.1 Constraint Extraction:

Figure (4) summarizes the constraints extraction process which can be classified into two major steps:

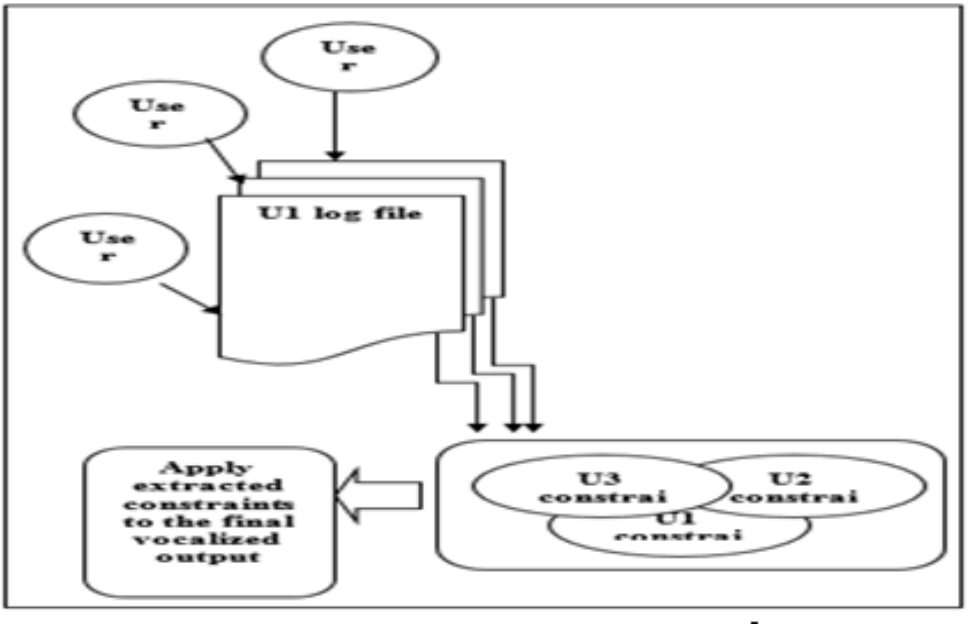

Figure 4: Constraints Extraction Process

6.2 Class Prediction:

The idea of predicting people's personality's stems from two basic ideas data collection and machine learning. For instance, some studies have shown that we can predict users' personalities from the pattern used by the user in social media such as Facebook or Twitter can be used to [9][10][11].

In our approach the data collection will be mainly from the users log file and set of predefined attributes if our queries limited to a known domain as we mentioned earlier in this section.

Since the relationship between extracted features might be in non-linear behavior we can apply PCA in order to select most relevant features to our query and user domain then classify each user as low, medium or high.

6.3 Mapping Constraints:

This step aims to map each category to appropriate set of constraints including the precision value tied to each attribute in the vocalized query.

7- Conclusion and future work

Current trends towards voice-based interfaces motivate the problem of data vocalization, a complementary problem to data visualization. In the proposed method gives a way to classify the users and their interest of how to select between the simple voice outputs row by row and concise version to find the good policies to select between simple and concise version in an interesting direction. This work is considered as extend to work proposed by by Trummer et al.[[1]. We are highly recommended extending this research in the future to be implemented in order to verifying and introduce proof of concept, with actual results.

8- References

1. Immanuel Trummer, Jiancheng Zhu and Mark Bryan, "Data Vocalization: Optimizing Voice Output of Relational Data", In VLDB Volume 10, Pages1574- 1585, 2017.
2. G. Lyons, V. Tran, C. Binnig, U. Cetintemel, and T. Kraska."Making the case for Query-by-Voice with EchoQuery". InSIGMOD, pages 2129–2132, 2016.
3. F. Li and H. Jagadish. "Understanding Natural Language Queries over Relational Databases ". SIGMOD Record, 45(1):6–13, 2016.
4. Fei Li, H.V. Jagadish, "NaLIR: An Interactive Natural Language Interface for Querying Relational Databases", SIGMOD'14, June 22–27, 2014
5. T.Hermann, A. Hunt, J.G. Neuhaoff, "The Sonification handbook", 2011.

6. Ryosuk Inoue, Yoshiaki Kurosawa, Kazuya Mera and Toshiyuki Taezawa, "A Question – And-Answer Classification Technique for Contracting and Managing Spoken Dialog System ", 978-1-4577-0931-9/11, IEEE, 2011.
7. Enc, J. Horvitz and David M. Pennock, "Methods And Apparatus For Predicting And Selectively Collecting Preferences Based On Personality Diagnosis", United States Patent, 2008.
8. Y.de Montoya*, J. Quoidbach*, F. Robic*, A.Pentland,"Predicting people personality using novel mobile phone-based metrics". Social Computing, Behavioral-Cultural Modeling and Prediction (2013).
9. Back, M.D., et al.: Facebook profiles reflect actual personality, not self-idealization. In: Psychological Science 21.3, pp.372-374 (2010).
10. Counts, S., Stecher, K.: Self-presentation of personality during online profile creation. In.: Proc. AAAI Conf. on Weblogs and Social Media (ICWSM). (2009) .
11. Stecher, K., Counts, S.: Spontaneous inference of personality traits and effects on memory for online profiles. In: Proc. Int. AAAI Conference on Weblogs and Social Media (ICWSM) (2008).

بيانات الصوت

تعزيز الصوت الناتج من قاعدة البيانات العلائقية

الملخص: أدت التطورات الحديثة في التعرف التلقائي على الكلام ومعالجة اللغات الطبيعية إلى جيل جديد من الواجهات الصوتية القوية. ومع ذلك، هناك القليل جداً من العمل على استخدام الواجهات المستندة إلى الصوت؛ للاستعلام عن أنظمة قواعد البيانات. وتحفز الاتجاهات الحالية نحو السطوح البينية القائمة على الصوت مشكلة غناء البيانات، وهي مشكلة تكميلية لتصور البيانات. لذا، يهدف البحث في رؤية البيانات إلى إيجاد أفضل طريقة لتقديم البيانات عبر واجهات مرئية. في هذه الورقة سوف نناقش طريقة لتصنيف المستخدمين واهتمامهم بكيفية الاختيار بين المخرجات الصوتية البسيطة الصف حسب الصف والصيغة المختصرة للعثور على السياسات الجيدة للاختيار بين إصدار بسيط وموجز هو اتجاه مثير للاهتمام

الكلمات المفتاحية : بيانات الصوت، قاعدة البيانات العلائقية ، تخصيص وقت التحدث.