April 2020

Matching Search Engine Responses to User's Domain Expertise Level

Marcos Calvo

Jakub Kriz

Jakob Zwiener

Matching Search Engine Responses to User's Domain Expertise Level

ABSTRACT

A user's query to a search engine or virtual assistant may reasonably be answered at several levels of complexity. In current search or information-retrieval engines, the user's domain expertise is not typically accounted for when providing a response. Thus, even if the answer is correct, it may not be appropriate for the user given their prior knowledge about the topic. This disclosure describes techniques that, with user permission, determine a user's domain expertise and select an answer appropriate to the user's level in the domain of the query.

KEYWORDS

- Domain expertise
- Subject matter expertise
- Content level
- Content appropriateness
- User profile
- Search engine
- Information retrieval
- Virtual assistant

BACKGROUND

A user's query to a search engine or virtual assistant may reasonably be answered at several levels of complexity, e.g., novice, amateur, college-level, power-user, researcher, expert, etc. In current search or information-retrieval engines, the user's domain expertise is not typically accounted for when providing a response. Thus, even if the answer is correct, it may not be appropriate for the user. An answer to a query such as "who were the Romans" may

contain too little detail (“a historic empire”) for a professor of history, or too much detail (“an empire centered on the city of Rome, present-day Italy, that arose (legendarily) around 753 BC, … ancient Roman politics, socio-economics …”) for a beginning student of history.

In general, quick and simple explanations are appropriate for novices in the domain of the query, while full, nuanced, or complex answers, based on specialized content, e.g., domain-specific websites, knowledge bases, etc., are appropriate for a professional or researcher experienced in the domain of the query.

Current content-filtering techniques, based on the known or stated age of the user, typically check for age-appropriateness of content. Such techniques do not extend to the user’s domain knowledge. Further, such techniques restrict the returned results rather than match or adapt results to the user’s prior domain knowledge.

Content-disambiguation techniques sometimes model users to determine the correctness of an answer to an ambiguous question. For example, if a user is identified as a programmer, a query about languages may result in answers related to programming; if the user is modeled as a linguist, the same query may result in answers related to linguistics. However, such techniques do not select a particular answer from equally correct answers of differing technical levels.

DESCRIPTION

This disclosure describes techniques to answer a user’s questions in a personalized manner appropriate to user interests and level of domain knowledge, determined with user permission, and to improve such answers.

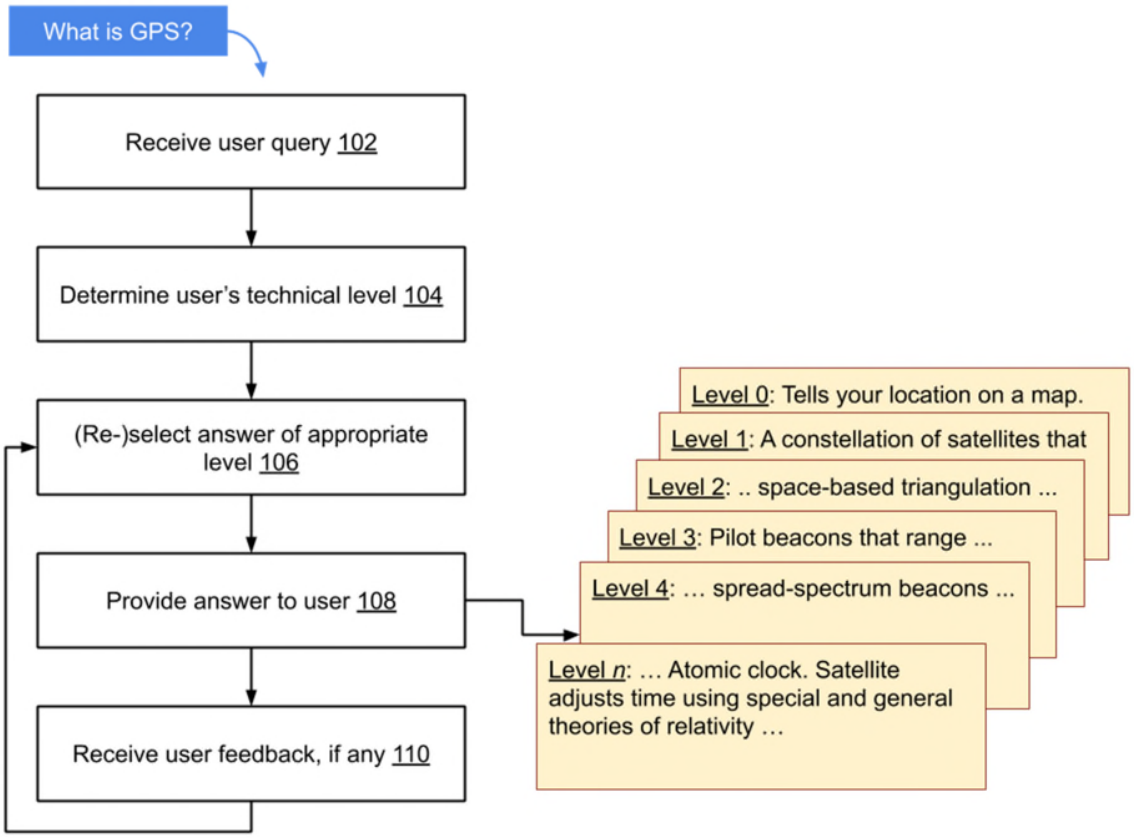

Fig. 1: Matching search-engine responses to user's domain expertise

Fig. 1 illustrates matching the responses of a search-engine, virtual assistant, or other question-answering or information retrieval system to a user's domain expertise, per the techniques of this disclosure. A user query is received (102), for example, "what is GPS?". With user permission, the user's technical level in the domain of the query is determined (104) using techniques described in greater detail below. An answer of appropriate level is selected (106) out of several possible correct answers (or websites) of differing technical levels. Techniques for selecting answers are described in greater detail below. The answer is provided to the user (108). The answer can be of a level that is appropriate for the user. As illustrated in Fig. 1, *n* levels of answers, e.g., a basic level answer of level 0 and progressively more technical/in-depth answers

at higher levels are available. The response that is provided to a particular user is selected based on the level appropriate for the user.

The user interface includes mechanisms to provide feedback, e.g., buttons labeled as "the answer is too simple," "the answer is too complex," etc. If the user provides feedback (110) by clicking on a feedback button, the answer is re-selected based on the user feedback. In this manner, a user can refine the content and language of the answer until it is commensurate with the user's level of domain understanding. This feedback can also be used to reassess the estimated user competence in the topic and to select better suited answers in the future.

Determining the user's level of domain expertise

The user's level of domain expertise can be determined in various ways. For example, the user can be provided with a one-time form or questionnaire that records the user's interests, experience in various fields, education level, etc. Alternatively, with user permission, the user's activity, e.g., browsing history or other online activity, can be analyzed to determine the user's domain expertise and/or interests. If the user does not provide input via the questionnaire and does not provide permission to access data that can be used determine user's domain expertise, the answer selection can proceed without such information, e.g., provide an answer that is suitable for all audiences, and the user can be provided options to request answer refinements.

For example, if the user browses websites that include popular workout content that uses easy-to-understand terms, then it can be reasonably inferred that the user is likely not a physiotherapist or fitness instructor. On the other hand, if the user routinely consults online resources that include human physiology and the science of exercise and utilize terms that require considerable prerequisite knowledge to understand, then it can be reasonably inferred

that the user has substantial knowledge in the domain, like a physiotherapist, fitness instructor, or other fitness professional.

As another example, if the user is a member of a social network group on a specialized topic, e.g., automotive history, then it can be reasonably inferred that the user has substantial knowledge of the specialized topic as well as of allied topics, e.g., automobiles, history, etc. Further, the level of user activity within the social-network group, e.g., passive member, occasional commenter, frequent commenter, number of followers, etc., is an indicator of the user's expertise on the topic of the social-network group.

In this manner, analysis of the user's online activities performed with user permission, can yield features of the user's profile, such as the user's age bucket, interests, domain knowledge, etc. Machine learning techniques can be used to classify users into different domain-based experience buckets. To track a user's changing level of domain expertise or knowledge, the user's profile can be updated periodically. Once created or updated, the user profile, e.g., the user's domain knowledge, is usable in the formulation or selection of an answer to a user's query.

Selecting answers at a level appropriate to the user's domain knowledge

To select answers at a level appropriate to the user's domain knowledge, information retrieval or question-answering systems, which can utilize machine-learning models, can be trained with labeled data that include not only correct answers to training queries but also user profiles that match a given correct answer.

Answers can be classified based on technical level by using language analysis techniques. For example, a greater frequency of certain words or phrases in an answer (relative to the frequency of appearance of those words in a standard corpus of documents) can indicate an

answer of higher technical level. Numerical measures such as term frequency inverse document frequency (TF-IDF) can be used to determine or characterize the technical level of an answer.

Answers provided by the question-answering system can be adapted, e.g., by use of natural language processing techniques, to match the user's skill level. For example, for novice-level users, complex content can be summarized, or the language of the answer simplified, using text-summarization techniques. For expert users, the question-answering system can search for further documents, e.g., of greater technical level, related to an initially provided answer.

In this manner, aside from providing correct answers to user queries, the techniques of this disclosure answer a user's question in a manner that is suitable to the level of the user's expertise in the domain of the question. The techniques improve the usefulness of the answer provided to the question on hand and enhance the usefulness of the question-answering system.

Further to the descriptions above, a user may be provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein may enable collection of user information (e.g., information about a user's social network, social actions or activities, browsing history, profession, a user's preferences, or a user's current location), and if the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used, so that personally identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level), so that a particular location of a user cannot be determined. Thus, the user may have control over what information is collected about the user, how that information is used, and what information is provided to the user.

CONCLUSION

This disclosure describes techniques that, with user permission, determine a user's domain expertise and select an answer appropriate to the user's level in the domain of the query.