

The Information-Seeking Problem in Human-Technology Interaction

Mohammad Alsulami

College of Computer Science and Information Systems,
Najran University,
Najran, Saudi Arabia

Asadullah Shaikh

College of Computer Science and Information Systems,
Najran University,
Najran, Saudi Arabia

and

Faculty of Computer Science and Information Technology,
Institute of Business and Technology (IBT),
Karachi, Pakistan

Abstract—In the history of information-seeking, the intention of a query and the posed query have some level of distance between them. Because human query-responders are innately connected to times and trends and have the ability to understand natural language and human intention, they have often been the idealistic sources of knowledge-direction. As the quantity of depth of knowledge of humanity grows, technological systems have sought to utilize natural language, both spoken and written, as a format of accepted queries. Modern works seek to improve such systems utilizing distance-metrics of literal queries to understood questions with maps to knowledge-bases. However, these methods do not often take into account the value of information in terms of query interpretation for mapping and as such may have identifiable limitations compared with human responders. In this paper, a model for information value is proposed and existing works in speech and query recognition are discussed relative to their considerations of information value.

Keywords—Information seeking; seeking problem in HCI; HCI Information seeking.

I. INTRODUCTION

Gathering information has been an integral human task throughout all of history and the modern world is no exception to this. When asked the answer to a particular query, someone being interviewed may be able to venture an answer. However, in order for a technological system to respond to such a query, there are many factors that must be considered that might almost be automatically taken into account by a human responder. For example, the coherency and legitimacy of a question may be determinations that happen quite unconsciously. Even corrections to queries that may be interpreted as incoherent and illegitimate may be corrected. Such corrections are hereafter often referred to as interpretations because that is exactly what they are. All human interaction and response in way of communication of information can be seen as a transmission and interpretation process. Perhaps what is most important to consider when identifying the nature of a question is what an information-seeker intended to ask or, rather, what he intended to find.

As human communication in the form of querying is most logically-performed in the medium from which questions initially arise, that of natural language, it seems clear that the most direct and sound means of making and interpreting queries

would happen in the form of natural language. However, natural language of humans tends to take two forms with almost equally-difficult challenges: spoken and written. Spoken language in the form of interpreted sound waves operates on such a challenging level that most technological applications choose to transform it into written language, which can be handled more compactly with specific construction. One of the major challenges in handling this aspect of human communication is that of correctly “hearing” what someone has said. Human capacity of correcting, distorting, and exaggerating what has been said is so profound that most people have heard entire sentences that were never actually stated! As such, it may be able to be stated that we as humans respond to what we “think” we hear and not what was actually said. Additionally, when responding to queries from other people, many human beings will construct the query within their own minds and perform many evaluations, distortions, and interpretations, perhaps even going so far as to rephrase the query in an entirely different (and hopefully desirable) way. Thus, the human ability to interpret the intention of a spoken query is quite profound, though modern advances in voice recognition, treated herein, have made great steps at approaching such a level of ability in technological applications. These devices though, it cannot be ignored, still translate from speech to text, thus transitioning into the other domain of human communication.

Written communication is perhaps arguably the clearest and most readily-processed form of human communication. In fact, “with difficult messages, both persuasion and comprehension of persuasive material were found to be greater when the message was written, compared to videotaped or audiotaped [4]”. Even still, it cannot be ignored that intention may be lost if incorrect statement or interpretation of a query is made when searching for a particular piece of information. Herein the ways in which intention is attempted to be extracted from queries and the nature of information-seeking as it pertains to the value of sought-after information are discussed. Additional discussion includes understanding the gap that exists between a human responder and a technological responder in terms of function and ability and the way in which the “distance” between what has been requested and what is understood to have been requested. The latter factor has been addressed by many existing works in the domains of speech recognition and

search engine implementation.

However, many of these systems attempt to pigeonhole an information-seekers queries using question-answer (QA) databases and other forms of context-limitation in order to ensure that an information-seekers queries are translated into something believed to be meaningful by the creators of such knowledge-bases. This is the approach taken in [1], where “we use information on the internet in the QA domain to find all possible question patterns, then use it to correct the queries that are partially recognized by speech-recognition software [1]”. In this work, the value of information is considered as an extension to the simple aspect of quality of information and discussion is presented on how limiting systems and technologies attempt to produce value by distilling queries into a finite realm of knowledge, much like seeing a world full of nails when possessing a toolbox with only a hammer. Limiting approaches in interpreting queries, while pretending to be open oracles of knowledge, attempt to pigeonhole an information-seekers intended information-target into the knowledge-base of the oracle. Because the intention of a query lies within the information-seeker and the knowledge of humanity is the product of a history of communication, we first begin with a historical treatment of the history of searching and a background of the existing works in query-recognition.

II. BACKGROUND: A HISTORICAL PERSPECTIVE OF SEARCHING

In the world of modern technology, information is the currency. Today's world is about living and breathing vast amounts of information, ranging from simple data like emails to complex data such as answers to deep technical questions. It is obviously-clear that more searchable information exists now than two decades ago by a factor of many, many times. Perhaps at one point it might have been the task of researchers and even everyday people to put forth great effort in even attempting to locate the information that they sought. Perusing libraries, pouring over collections of statistics, seeking personal interviews, and even undertaking private investigations might have been necessary just to obtain information related to that which was being sought. At that time, finding too little information may have often been the norm. Those days are, for the most part, gone.

Today, one can access so much information with so little effort that the quantity can be mindboggling without even attempting to gather so much. One search of a popular search engine can yield so much data that the whole of the world's libraries fifty years ago would collectively seem like an abbreviated encyclopedia in comparison. This information overload can make even the simplest searching tasks much more complex. The problem now is how to reduce the information that we collect to only the pertinent information that is desired. Maybe finding movie times at one point required a short stroll or a phone call. Now, a simple search for movie times on a smart phone might lead to confusion about what which of the multitudes of theatre results are the ones that match our intended destination. With this switch in paradigm new problems have arisen while lingering factors from the bygone days of searches in information shortage still remain.

It might be surprising to some that freedom comes at a cost. This statement may also lead to the possibly equally-surprising

question: What kinds of freedom are there in searching for information? The answer to this question is quite simple: multitudes. Before delving into the academic nature of such freedom, it can be helpful to understand a brief history of what freedom means in terms of information-searching and why it has evolved with technology and culture. Fifty years ago, if one wanted the history of the rule of succession of British Empire, there were many choices available. The simplest choice might have been a trip to a reasonably-sized library, possibly of public, school, university, or private nature, and a consultation of a textbook of appropriate title. Perusing the index of such a textbook might or might not have yielded a reference to the information that was sought. It could and can be concluded that such a textbook was arranged with the intent on making the knowledge contained within reasonably complete and reasonably accessible.

However, it should also be observed that such assurances are based on the authors understandings of societal and cultural norms regarding the processes of searching for information as well as their own personal opinions on how information should be arranged and made available for search. Thus, searching multiple textbooks for the same knowledge might have led to similar ways of searching or perhaps completely different ways. As to the effectiveness of such a search, this vastly depended on ones own personal means and concepts of search compared with those arranging the information being sought. Perhaps the desired textbook was in the library, but finding this textbook may have required yet another search of a completely different nature, such as through a card catalog. This system may have been arranged in a completely different fashion than that of the organization of the desired textbook itself. If one knew the title of the textbook, finding it may have been relatively easy. If the title of the textbook was unknown or one did not know if such a textbook even existed and was attempting to determine the existence and location of such a textbook, this put the searcher at the mercy of the textbook-locating system. Of all textbook-locating systems, the one that is perhaps the closest to modern concerns was also often the most powerful: someone who knows, be they scholar, professor, or librarian.

Dating to ancient times, persons of knowledge have not only been the repositories of important knowledge, but also the means of seeking such knowledge. Wise-men, scholars, and teachers have ever played an important role in society. If they did not know the desired knowledge themselves, they could likely direct an information-seeker in an appropriate direction. Perhaps none was so famous a person in directing those to desired information than “the oracle” from which many modern information systems derive a basic term in describing their natures. One needed only ask the oracle and the desired information would be given or direction would be given that would reveal the desired information. Perhaps this “reveal rather than teach” approach allowed for a certain degree of information-isolation to the most appropriate and knowledgeable containers, a concept that also deserves academic treatment. Whatever the case, this human source of knowledge-direction had a crucial benefit over mechanical and artificial means of information search: the ability to speak ones own language.

It is hardly-debatable that many people of the past and of

the present would almost exclusively prefer to query another human being regarding where to find certain information being sought. Modern call centers paired with automated hotlines are clear examples of such preference. Perhaps there is a clear emotional component of human-human interaction in this preference. However, if asked for a reason for preferring to query another human being, the logical response that someone might simply give is that it is faster than using a technological system. Modern technological systems attempt to mimic this behavior by allowing natural language queries that attempt to understand what an information-seeker is intending by literal interpretation and cross-matching of query terms. Several of the methods proposed for improving modern query and informational systems are discussed herein.

One such method is that of the RSVP system proposed in [1] that builds a QA knowledge-base of query questions by collecting a very large set of existing questions from QA websites. The authors of such work admit that there is very low coverage of all possible questions and propose that questions can be “clustered” into patterns for 99% coverage of all possible types of questions. This is the means in which they believe that can improve speech-recognition software in the QA domain. The methodic approach taken in [1] is that of attempting to minimize the distance between an interpreted query and the intended query and create a similar minimization to the database of QA question structures available. Related to the attempt of determining differences between queries and intended queries is the work in [2] in which the authors show that “information distance is a universal cognitive similarity distance [2]” and state that “information distance between individual objects is needed in pattern recognition where one wants to express effective notions of ‘pattern similarity’ or ‘cognitive similarity’ [2]”. Many of the same authors of this work also attempted to verify information-evolution algorithms based on genetic concepts through analysis of chain-letters in [3].

This relates in how information interpretation and meaning can evolve with time in response to trend. The algorithm proposed by the authors of [3], while originally developed for genomes, was applied in the realm of languages for detecting plagiarism in student assignments. This concept relates to how queries can be posed in different ways and mapped to similar information. In similar means to [1], [5] proposed a plagiarism-detection algorithm based on “a fundamental question in information theory and in computer science” on “how to measure similarity or the amount of shared information between two sequences [5]”. For performing this measurement, the authors proposed a metric based on Kolmogorov complexity and claim to have proven it to be universal. This, again, relates to posed queries and intended information in that evolution of similar intention is analyzed for similarity with trend identified. In terms of search-engine-oriented distance differences as a means of mapping queries [6], presents a theory of similarity between words and phrases based on information distance and Kolmogorov complexity. The authors therein present applications in hierarchical clustering, classification, and language translation based on the use of search results from a popular search engine based on the information trends of millions of independent web-knowledge-providers. The method proposed by the authors does take into account trending of information as the web evolves based on the evolution of search engine

results from scouring the Internet over time.

In the work of [7], the authors attempt to correctly anchor in time and space a number of entities, implying a labelling system to reach such entities as factual objects. This may lead to certain prejudices of organization and labelling towards the designers ways of thinking. In contrast [8], proposes and tests a parameter-free algorithm that “would limit our ability to impose our prejudices, expectations, and presumptions on the problem at hand, and would let the data itself speak to us [8]”. The authors show that their approach is competitive or superior to state-of-the-art approaches in “anomaly/interestingness detection, classification, and clustering” utilizing empirical tests. This approach seems to allow for more independent expression of information-seeking that contradicts the more rigid means of many of the other works. The work of [9] relates to the continual seeking of proper metrics for information distance by addressing the fact that “the universality theorems for normalized information distances were only proved in a weak form.” The authors of this work expose some of the issues of information distance theory in attempting to establish a better means of measuring information distance. This is similar to the work of [10] in which the authors study “a new class of distances appropriate for measuring similarity relations between sequences, say one type of similarity per distance [10]”. In a completely different way [11], with its appropriate title of an incorrectly-interpreted voice-query “How to wreck a nice beach you sing calm incense”, takes the approach of solving the problem of speech-recognition systems inappropriately-ordering hypotheses for similar-sounding phrases. In their work, the authors proposed “a supplementary method for ordering hypotheses based on Commonsense Knowledge [11]”, the system they developed by “filtering acoustical and word-frequency hypotheses by testing their plausibility with a semantic network derived from 700,000 statements about everyday life.”

In response to these methods, herein a model is proposed of information value as a means of mapping queries to information. This model attempts to account for factors that may be overlooked by many of the literal-matching-and-correction schemes of current works.

III. METHODS: QUANTIFYING A SEARCH

While quality of information is typically paramount on the mind of the searcher, the gaining of such excellent information is ultimately determined useless if the time spent obtaining such information were to exceed the reasonable scope of time when such information would be considered valuable. For example, finding out the time of showing of a movie in a theatre an hour after the movie ends is worthless. This scope may be lessened by the information-seekers own personal preference, meaning that the value of the sought-after information may be eclipsed by other demands on the information-seekers time-resources. For example, an information-seeker may be attempting to review products while standing in a store and eventually forgo obtaining additional information, or any information at all, about products to be purchased if the information-seeker determines that the benefits of such additional information are not worth the effort of obtaining it. This viewpoint of information-seeking leads one to the mindset that what is of importance in obtaining information is a

matter of efficiency. Using the common definition of efficiency, referring to the ratio of output to input in a process, we can frame the efficiency of a process as

$$Efficiency = OutputGained/InputPresented.(1)$$

In considering the concept of efficiency, it is a basic economic fact that in prior centuries materials were extremely expensive compared with the cost of human time. Many important achievements of mankind have been historically achieved through raw effort of input of human time with most expense being considered in terms of the materials needed to reach such achievements. Thus, the price of food and travel may have been in the past of considerable more concern in the process of seeking of in-formation than the time spent by a person. The reverse of this paradigm has almost exactly corresponded with the paradigm reverse of information shortage versus overload. In the modern technological and business world, it is a known fact that fewer commodities are more expensive than human time. The correspondence of material-availability and information-availability is difficult to ignore. With so much information being available for ready-access, the definition of efficiency in information-seeking can be redefined as

$$Information - SeekingEfficiency = Quality/SearchTime.(2)$$

Considering that most human beings will likely not seek quality information that has subjectively-little value, what remains to be considered is how to determine the subjective value of sought-after information as it relates to quality. According to [12], the “value of search results as a whole, a utility measure, was found to be the best single measure of interactive information performance (success) among the 20 measures selected for study [12]”.

The quality of information can be considered constant relative to the actual information desired. One reason for this is because at the time a query is posed, the ideal answer exists pertaining to the time that the query was posed. Posing the same query later results in searching a different knowledge-base than existed at the previous time. Thus, at any given time, there exists a response to any query that could be considered ideal for such a query given the knowledge available at the time. The quality of the response to a query at a given time is relative to the knowledge available at that time. However, the value of a particular piece of information relative to a query could depend greatly upon the quality of the information given changes to the knowledge base. As such, it seems clear that quality of information, again always relative to a particular query, can change with the knowledge-base, which can change with time. Given that quality has been established for a particular query at a particular time, it can be presumed that the mapping of the query to the knowledge base may change with time, leading to a single query mapping to different pieces of information at different times. For example, a query for the current movies showing at a particular theatre would result in very different information for equal qualities at different times. Thus, there must be a component of information value that relates quality to time based on the information-seeker and his particular query, a kind of pertinence. For example, the best way to treat a particular medical condition might be a piece of information of particular quality with value that changes with time based on the information-seeker that would warrant

a level of pertinence. Even pertinence, however, is not enough to determine whether information has value at a particular time.

Even a pertinent piece of information may not be of significant value if it is not what is actually desired. The trend of information can vary personally and societally with time. A particular piece of information may be highly-valued during a particular scope of time simply due to its popularity as a trend. Popular search results on major search engines make good examples of this. Thus, when establishing the value of a piece of information, the trend of simple human interest must be taken into account. Tracing this trend can lead to another piece of information in itself, which is how preference for information evolves with time. Sometimes the value of information depends heavily upon the specific time that the information is being sought. Five days before a movie is to show at a theatre, the information regarding show times for that theatre may be of some value to the collective audience that will eventually see the shows on the destined day. This information will likely be of increasing value as the destined day approaches and then be of basically no value the next day. This means that the trend of information can grow as a time approaches and then fall to nothing shortly after that time has been reached. Under these premises, the definition of information value could be stated as

$$InformationValue = Quality*Pertinence*Trend.(3)$$

The definition in (3) establishes a model for information value though it does not yet determine what exactly constitutes quality of information. Basic academic thought might lead one to understand that what is normally intended with quality is correct information that is of significant enough quantity. In terms of the value of this quality of information, the pertinence and trend of information should act to determine its value. For example, an academic article or textbook entry might be of the highest quality for a particular query though an information-seeker might not desire this quality of information and thus a trend might be established towards a particular lower quality information that is either not as correct, not as voluminous, or both. This could explain the popularity of Internet search engine results and online encyclopedia articles being sought after. A particular model for information quality could be framed as

$$InformationQuality = Correctness * Quantity.(4)$$

Combining definitions (2), (3), and (4) yields a more complex model for framing information value once rearranged as

$$InformationValue = Correctness * Quantity * Pertinence * Trend.(5)$$

The definition in (5) still does not seem to take into account the value of information as it relates to efficiency. Even given the availability of correct information of adequate quantity that is pertinent to the information-seekers query that has an established trend of desire, an information-seeker may still forgo obtaining such information. Returning to (3) and substituting efficiency for quality, (5) can be rewritten as

$$InformationValue = Correctness * Quantity * Pertinence * Trend/SearchTime.(6)$$

With so many of the terms in (6) depending on time, society, and personal consideration, one might be led to wonder

how an information-seeker could possibly find the information being sought. The authors of [7] attempted to address this with a project undertaking based on ensuring that “entities, facts, and events are anchored in both time and space [7]”. Perhaps there are some simple facts to consider in this. Historically, much of the information available at any given time has been through the devices, means, and trends of the day. This means that the pertinence and trend of the information are established within the resources of the day. Likewise, the information on how to find information was established by trend such that the correctness and quantity of such information to find other information was obtainable with reasonable search time. This could have occurred through how the people of the day were taught. By establishing the pertinence of such information-seeking sources, it may be that the value of information through history could have been carefully controlled and monitored on a societal level. In the modern world of search engines and QA databases, data tend to live forever, meaning that the evolution of data in the form of availability as a function of decay and growth relative to time is distorted.

In the modern Internet world, querying for information has been revolutionized and evolved by the sciences behind search engines and QA databases. The widespread, growing use of information in almost every aspect of life and business has driven these developments. As might seem a natural evolution, information-seeking technology has evolved towards natural-language systems allowing for queries in spoken and written form. These systems typically operate under the presumption that an information-seeker understands exactly the information he is seeking and that his query is worded to be directly-mapped to the intended information. This approach does not take into account many aspects of value, which will be discussed later herein. Although technology has managed to preserve and even widen the availability of data that would have previously decayed or become unavailable, such as recorded music and records, the value of data continues to change with society and the times. Modern information-seekers can find records, information, and cultural data from fifty years ago, which is an enormous contrast from the availability of one-hundred-year-old information at a point in time fifty years ago. What remains to be discussed is how modern query-response systems deal with this fact by attempting to understand the right questions and how those systems do or do not take into account pertinence and trend of information.

IV. EVALUATION: HAVING THE RIGHT QUESTIONS UNDERSTOOD

Returning to the concept that an information-seeker may have many ways of attempting to obtain information, considering pertinence and trend to have been established by the information-seeker and utilizing (6) leads to the conclusion that there is a balance between the quality of information being sought and the search time. The pertinence and trend of the information may be considered as a kind of filter to limit the information-seekers search while the efficiency of search drives the information-seekers actions. However, the medium of querying for information, such as a library in the past or the Internet in the present, may not have been established with clear defining lines to allow pertinence and trend to be taken into account. This may lead to a misunderstanding of what the information-seeker is asking that might be obvious to a

human oracle whose knowledge-base of information-seeking is established in the times of the information-seeker. One clear area of difference between an Internet search and an interview-search is the recognition of language in queries.

A human information-seekers own knowledge of how to seek information is firmly rooted in language, which is firmly rooted in the times. Thus, the diction and tendencies of the information-seeker may well be heavily-affected by the society and age in which he finds himself. This seems to lead to a point that language should be considered as a kind of knowledge-base that itself is dictated by information values. When considering the means of querying for information, human language is a natural choice as it is the basis in which thought and writing are established and likely the principally-understood form of most original queries. Considering that a human oracle whom an information-seeker may query has established values in language based on his upbringing and societal interaction, a human oracle may be able to interpret an information-seekers query utilizing his own value system of language. Whether reading a query or hearing it spoken, a human oracle has an advantage over a designed or programmed one: the ability to interpret a query based on his own querying ability.

When designing an oracle, a big challenge is one that changes its interpretations of queries with the values of information-seekers. As a programmed oracle, such as a search engine, does not inherently have the ability to establish itself in the times and determine its own values of information, it must rely on being given this information or deriving it. A major search engine may be able to notice trends in querying and information quite readily. Perhaps what the search engine may lack is the ability to understand the motivations behind a query. These motivations are likely established within the times and society of the information-seeker. Even a query of history comes from the present time. When communicating with a search engine, the interpretation of a query needs to be based on what the information-seeker intended to find. This can be rather difficult on several fronts of understanding.

When human information-seekers search for information using a search engine or other search technology, they often do not utilize the same protocols as they would if querying other humans. As such, queries for information posed to technological systems are often incorrect grammatically and structurally in a way that might be considered insulting to a human responder. An example might be “famous action movie 1966”, which sounds more like a trivia statement than a query. One might be led to believe that the intended information here is the name of a particular famous action movie released in 1966, however the intention to this apparent clarification question is somewhat vague. Likewise, a query of “blue soccer ball” is more of a statement than a query and thus the intention is very unclear. However, it seems from the general usage of search engines in modern times that information-seekers have adapted to the way in which search engines operate and intentionally pose such queries with a belief that certain information will be returned. For modern speech-recognition and QA query-matching systems, this mode of operation can be potentially defeating of natural language processing and correction methods that they attempt to utilize. Several such current methods are herein discussed and analyzed.

The RSVP system proposed in [1] that builds a QA knowledge-base of query questions by collecting a very large set of existing questions from QA websites attempts to establish value by assuming a very broad range of possible questions that an information-seeker may ask and then literally-correcting queries to fit these questions. This is a bit like trying to record the thoughts of millions of people in order to learn a language. While trends of information may be captured if they evolving QA knowledge-bases are consistently kept in-sync with the times, the pertinence to individual queries is potentially lost by pigeonholing the information-seekers ways of posing queries. Additionally, this type of system can greatly suffer from ambiguously-posed queries that lack substantial qualification. The proposed system does attempt to equate related means of posing a query about a single piece of knowledge though perhaps in a crude, collective way. The statement of the authors in [2] that “information distance is a universal cognitive similarity distance [2]” as the basis of their efforts fails to take into account trends and localized pertinence of information. An example of such might be the profanities of one culture being typical speech in another. In their related work in [3] based on evolution of meaning and communication in chain-letter-generations, the authors do account for trending over time.

Thus, the intention of posed queries of similar semantic nature may evolve as time progresses, much as was noticed in the analyzed chain-letters. Works [1] and [5] used the application of plagiarism-detection in determining an appropriate way to measure similarity and shared information between two sequences of information. This literal interpretation of distance rather than a value-based interpretation of distance may be able to correct and identify certain semantic variations and facilitate a many-to-one mapping of knowledge with a certain degree of potential trending, but such as an approach may also suffer in being able to distinguish pertinence of information requests that are distinct in some way. The more vaguely a query is posed, the less effectively this method will likely be to discern subtle differences. The authors of these works proposed a metric based on Kolmogorov complexity and claim to have proven it to be universal. This, again, does not account for localized and temporal differences in trend and pertinence.

These approaches lead to the efforts of [6] regarding search-engine distances for mapping queries to information. The authors here use search results from a popular search engine based on the information trends of millions of independent web-knowledge-providers, which does in fact take into account trending of information as the web evolves based on the evolution of search engine results from scouring the Internet over time. This approach is unique in that it takes into account the ever-general population of online-user culture in regards to information-seeking. As an information-seeker is likely also an avid Internet user in modern times, these two cultures are likely closely-enough synchronized to allow for generality of trend and pertinence for time and space to be taken into account in interpreting an information-seekers query. This allows the establishment of value of information pertaining to a query given that enough web contributors make the associated mappings and intention-relationships. In a similar, but more rigidly-structured approach, the authors of [7] attempted to correctly anchor in time and space a number of entities, implying a labelling system to reach such

entities as factual objects. This may lead to certain prejudices of organization and labelling towards the designers ways of thinking.

Additionally, such a controlled system that is not allowed to vary with trends and localized pertinence may be unable to maintain validity if significant deviation in knowledge-values occur on a broad enough spectrum. The approach of [7] is contrasted with that of [8] that proposes and tests a parameter-free algorithm that “would limit our ability to impose our prejudices, expectations, and presumptions on the problem at hand, and would let the data itself speak to us [8]”. The authors show that their approach is competitive or superior to state-of-the-art approaches in “anomaly/interestingness detection, classification, and clustering” utilizing empirical tests. The effort here in allowing data to establish its own trend and pertinence is appropriate in terms of establishing value as it separates the labelling of searching for information from the underlying knowledge and allows for free-form association without utilizing the authors own personal knowledge-mapping concepts in the method. This approach seems to allow for more independent expression of information-seeking that contradicts the more rigid means of many of the other works. In returning to literal, distance-oriented metrics, [9] sought to establish proper metrics for information distances. The authors of this work expose some of the issues of information distance theory, but do not make direct conclusion of the need to establish underlying value of information in terms of determining information relationships.

The authors of [10] took similar direction in studying a new class of distances for measuring similarity. This work makes progress in literal-interpretations and has many of the same shortcomings of other related works. In contrast to the many literal methods analyzed, the authors of [11] take a direct, value-oriented approach to correcting voice-recognition queries by strongly sorting hypothesized query meanings and matching them against a system appropriately named Commonsense Knowledge. This designers of this method strived to eliminate the need for information-seekers to self-correct their own queries, creating more seamless and believable voice-only searching capability. What is particularly interesting about this approach is the somewhat indirect means of establishing information value by determining if a query “makes sense”. This is interesting because a query “making sense” implies a general and localized understanding of information trend and pertinence. The fact that the authors used an extensive, temporally-oriented database of highly-pertinent phrases from “everyday life” inherently makes this system the strongest of those reviewed in terms of value-orientation of information-seeking.

V. DISCUSSIONS

In the modern world of information-seeking, technological methods have evolved considerably from the indexes, libraries, and interviewing methods of old. Modern information-seekers have many more degrees of freedom in regards to how they interact with searching technology leading to many advantages and disadvantages. The trend of information-locating systems towards accepting queries in natural language forms close to the thoughts driving such queries of information-seekers has allowed faster, more-direct, and more-specific searching

processes than the statically-arranged information-searching systems of history. Given the increasing excessive abundance of information, these processes are as much about filtering-out unwanted information as locating the wanted information. As such, it has become increasingly-important to understand the intended information that an information-seeker desires rather than assuming that by literally-interpreting a query. This is a matter of establishing the value of information to an information-seeker relative to the provided query and any other information known about the context of the user. Correspondingly many modern voice-recognition-based searching methods have been oriented to perform corrections based on established bounds for context in order to limit the degrees of freedom that a potentially-incorrect query may have to establish the appropriate query mapping to the information of value being sought.

This speech-to-text conversion has a similar counterpart in text-to-text conversion where calculated assumptions are made based on the context of the information-seekers relative location in the global knowledge-base. However, this attempt at arbitrarily-confining the pertinence and trend of sought-after information can inappropriately limit the degrees of freedom available to a user through the querying means. Establishing the correctness of information to a query by ensuring literal map-ping of query terms to a set of previously-mapped questions that supposedly have pertinence and trend toward information as a means of establishing value of information relative to a query assumes that the information-seeker's intentions and means of querying are similar to those used to derive the underlying mappings. Additionally, this method is complicated by the way in which information-seekers will pose queries to technological systems compared with similar query situations of actual humans.

When an information-seeker poses a question to another human, an established protocol requiring a certain, reasonable construction of the query as a question or command has already been instilled in both information-seeker and query-responder. Because the responder has an understanding of construction of queries and the ability to sense the supposed intention of the information-seeker in terms of the value of the information being sought relative to the information-seeker, the human responder can make corrections to misheard or misstated queries and even completely restate the query in order to arrive at the intended informational response. Technological systems do have not the luxury of being treated with human-human protocols and as such are often queried with incomplete, incoherent phrases that are often specifically-tailored to result in certain responses from technological systems and therefore may circumvent the ability of such systems to provide potentially-improved responses based upon complete and properly-constructed queries per human-human communication standards. Current technological QA systems lack the breadth of deriving intention from posed queries and instead rely upon literal interpretations of spoken or written queries with corrections being made by means of cross-examining existing databases of posed queries and mapping such queries to potentially-valuable information based on correct-query similarity. In this work, the current means of performing such context-limiting searches were reviewed with the perspective given of their attempts at providing informational-value through presuming pertinence and trend

based on the context and content of posed queries and focusing efforts on correctness of mapping to information through creative, literal, mathematical interpretations of query content. The approach at identifying how different "what was said" was from "what it heard" seems at times in current technology to be identifying "what is your purpose here". Perhaps this is the price that is paid for specialization of knowledge-bases and a lack of understanding in technological devices of human intention. Two of the systems reviewed that allowed free-form interpretations of information based on evolving trends in daily life were perhaps the best oriented at maintaining information value in query responses through a loose connection between information sought and the means of querying. Whatever the case may be, great advancements have been made in natural language searching technologies, be they spoken or written, though there is thankfully still "Room" for improvement.

VI. CONCLUSIONS

In this paper, we have proposed, a model for information value and existing works in speech and query recognition for information value. In practical, information-seeking, the purpose of a query and the posed query has some level of distance. We have addressed the value of information in terms of query interpretation for mapping and as it may vary in limitations compared with human responders.

REFERENCES

- [1] Yang Tang, Di Wang, Jing Bai, Xiaoyan Zhu, and Ming Li. Information distance between what I said and what it heard. *Commun. ACM* 56 (July 2013), 70-77.
- [2] Bennett, C.H., Gcs, P., Li, M., Vitnyi, P., and Zurek, W. Information distance. *IEEE Transactions on Information Theory* 44, 4 (July, 1998), 14071423.
- [3] Bennett, C.H., Li, M., and Ma, B. Chain letters and evolutionary histories. *Scientific American* 288, 6 (June 2003), 7681.
- [4] Chaiken, Shelly; Eagly, Alice H. Communication modality as a determinant of message persuasiveness and message comprehensibility. *Journal of Personality and Social Psychology*, Vol 34(4) (October 1976), 605-614.
- [5] Chen, X., Francia, B., Li, M., McKinnon, B., and Seker, A. Shared information and program plagiarism detection. *IEEE Transactions on Information Theory* 50, 7 (July 2004), 15451550.
- [6] Cilibrasi, R. and Vitnyi, P. The Google similarity distance. *IEEE Transactions on Knowledge and Data Engineering* 19, 3 (March 2007), 370383.
- [7] Hoffart, J., Suchanek, F.M., Berberich, K., and Weikum, G. YAGO2: A spatially and tempo-rally enhanced knowledgebase from Wikipedia. *Artificial Intelligence* 194 (January 2013), 2861.
- [8] Keogh, E., Lonardi, S., and Ratanamahatana, C.A. Towards parameter-free data mining. in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM Press, New York, 2004, 206215.
- [9] Li, M. Information distance and its applications. *International Journal on the Foundations of Computer Science* 18, 4 (August 2007), 669681.
- [10] Li, M., Chen, X., Li, X., Ma, B., and Vitnyi, P. The similarity metric. *IEEE Transactions on Information Theory* 50, 12 (December 2004), 32503264.
- [11] Lieberman, H., Faaborg, A., Daher, W., and Espinosa, J. How to wreck a nice beach you sing calm incense. in *Proceedings of the 10th International Conference on Intelligent User Interfaces (2005)*, 278280.
- [12] Su, Louise T Su. Value of search results as a whole as the best single measure of information retrieval performance. in *Information Processing & Management*, Volume 34, Issue 5, September 1998, 557-579.