J Intell Inf Syst (2007) 29:305–327
DOI 10.1007/s10844-006-0015-y

Context recognition using internet as a knowledge base

Aviv Segev · Moshe Leshno · Moshe Zviran

Received: 1 July 2004 / Revised: 26 November 2005 / Accepted: 30 January 2006 /
Published online: 24 January 2007

Abstract Context recognition is an important component of the common sense knowledge problem, which is one of the key research areas in the field of Artificial Intelligence. The paper develops a model of context recognition using the Internet as a knowledge base. The use of the Internet as a database for context recognition gives a context recognition model immediate access to a nearly infinite amount of data in a multiplicity of fields. Context is represented here as any textual description that is most commonly selected by a set of subjects to describe a given situation. The model input is based on any aspect of the situation that can be translated into text (such as: voice recognition, image recognition, facial expression interpretation, and smell identification). The research model is based on the streaming in text format of information that represents situations—Internet chats, e-mails, Shakespeare plays, or article abstracts. The comparison of the results of the algorithm with the results of human subjects yielded a very high agreement and correlation. The results showed there was no significant difference in the determination of context between the algorithm and the human subjects.

Keywords Record classification · Retrieval models · Metadata · Information filtering · Text analysis · Knowledge retrieval

1 Introduction

The question of context recognition is defined as one of the main questions addressed by the international interdisciplinary context community (AAAI, 1999). The effective recognition of context is essential in the common sense knowledge problem, considered to be one of the primary research areas in Artificial Intelligence (AI) (Ein-Dor, 1999).

A. Segev (✉)
Faculty of Industrial Engineering and Management, Technion, Haifa 32000, Israel
e-mail: asegev@technion.ac.il

M. Leshno · M. Zviran
Faculty of Management, Tel Aviv University, Tel Aviv 69978, Israel

Virtually every application of AI requires the use of context. At the very least, the application domain itself provides an important context for reasoning and AI applications can be more effective and efficient if they utilize this contextual knowledge. Frequently, the application must function in several different contexts. For example, an interface has to work with different users, a medical diagnosis system has to address different patients and different disease presentations, and an autonomous vehicle needs to work effectively in different geographical locales and with different terrain types. For applications to take their context into account and to profit from contextual knowledge, context and contextual knowledge must be represented explicitly. Many existing approaches in AI, both theoretical and applied, take the context into consideration implicitly. In other words, contextual knowledge may be present in the approach, but it is not explicitly identified as distinct from other kinds of knowledge. Furthermore, the context is not identified as something about which inferences can be made. This representation, like any other sort of implicit representation of knowledge, denies the reasoner access to its own knowledge about context (e.g., for learning or reasoning about its adequacy), leads to redundant representation, and makes it difficult for people to maintain the knowledge base (AAAI, 1999).

The rest of this section is organized as follows. Section 1.1 presents related works on context. Section 1.2 surveys research studies on the problem of information seeking and information retrieval. Section 1.3 presents works on context recognition. Section 1.4 presents an outline of the remainder of this paper.

1.1 Related works on context

Context has been researched from the aspects of artificial intelligence, natural languages, conversations, formalism of knowledge, goal planning, human expertise in context, knowledge representation, and expert systems.

McCarthy (1987) in his paper *Generality in Artificial Intelligence* mentioned some of the main problems existing in the field. The formalization of the notion of context was defined as one of the main problems. McCarthy argues that a most general context does not exist.

Consequently, the formalization of context and a formal theory of introducing context as formal objects were developed (McCarthy & Buvac, 1997). Context was introduced as abstract mathematical entities with properties useful in artificial intelligence. The context abstract definition was developed in the Cyc project in the form of microtheories (Guha, 1991). The formal theory of context was used to resolve lexical ambiguity and reason about disambiguation (Buvac, 1996).

The blackboard model of problem solving arose from the Hearsay speech understanding systems (Erman, Hayes-Roth, Lesser, & Reddy, 1980). These ideas were then extended into the standard blackboard architecture in Hearsay-II. The blackboard model has proven to be popular for AI problems and in the years since HS-II a variety of blackboard-based systems have been developed. HS-III was developed to integrate alternative representations. HS-III had a context mechanism that allowed the integration of knowledge to resolve uncertainty. Blackboard architectures have been used for interpretation problems such as speech understanding (Lesser, Fennell, Erman, & Reddy, 1975), signal understanding (Carver & Lesser, 1992), and image understanding (Williams, Lowrance, Hanson, & Riseman, 1977) and for planning and control (Hayes-Roth, 1985).

Blackboard architecture will be implemented in the context recognition model. The different attributes of the current "world state" are translated into text and added in turn to the blackboard. The data represented in the blackboard model serve as the input to the context recognition algorithm.

1.2 Information seeking and information retrieval

Information seeking is the process in which people turn to information resources in order to increase their level of knowledge in regards to their goals (Modica, Gal, & Jamil, 2001). Information seeking has influenced the way modern libraries operate (using instruments such as catalogs, classifications, and indexing) and has affected the World Wide Web in the form of search engines.

Although the basic concept of information seeking remains unchanged, the growing need for the automation of the process has called for innovative tools to assign some of the tasks involved in information seeking to the machine level. Thus, databases are extensively used for the efficient storage and retrieval of information. In addition, over the years techniques from the realm of Information Retrieval (Salton & McGill, 1983) were refined to predict the relevance of information to a person's needs and to identify appropriate information for a person to interact with. Finally, the use of computer-based ontologies (Smith & Poulter, 1999) was proposed to classify the available information based on some natural classification scheme that would permit more focused information seeking.

Valdes-Perez and Pereira (2000) developed an algorithm based on the concise all pairs profiling (CAPP) clustering method. This method approximates profiling of large classifications. Use of hierarchical structure was explored for classifying a large, heterogeneous collection of web content (Dumais & Chen, 2000). Another method involves checking the frequency of the possible keyphrases of articles using the Internet (Turney, 2002). However, this method is based on an existing set of keywords and uses the Internet for ranking purposes only.

There is an extensive body of literature and practice in the area of information science on ontology construction using tools such as a thesaurus (Aitchison, Gilchrist, & Bawden, 1997) and on terminology rationalization (Soergel, 1985) and matching of different ontologies (Schuyler, Hole, & Tuttle, 1993). In the area of databases and information systems many models were proposed to support the process of semantic reconciliation, including the SIMS project (Arens, Knoblock, & Shen, 1996), SCOPES (Ouksel & Naiman, 1994), dynamic classificational ontologies (Kahng & McLeod, 1996), COIN (Moulton, Madnick, & Siegel, 1998), and CoopWARE (Gal, 1999), to name a few. The ontology constructions can be seen as a manual effort to define relations between concepts, while context recognition attempts to identify, in our case automatically, instances of a given situation that could be related to a concept or concepts in the ontology framework.

This model attempts to automate context recognition, based on Information Seeking and Information Retrieval techniques, using the Internet as a database for possible contexts.

1.3 Context recognition

One context recognition approach addressed the creation of taxonomies from metadata (in XML/RDF) containing descriptions of learning resources (Papatheodorou, Vassiliou, & Simon, 2002). Following the application of basic text normalization techniques, an index was built, observed as a graph with learning resources as nodes connected by arcs labeled by the index words common to their metadata files. A cluster mining algorithm is applied to this graph and then the controlled vocabulary is selected statistically. However, a manual effort is necessary to organize the resulting clusters into hierarchies. When dealing with medium-sized corpora (a few hundred thousand words), the terminological network is too vast for manual analysis, and it is necessary to use data analysis tools for processing. Therefore, Assadi (1998) employed a clustering tool that utilizes specialized data analysis functions and clustered the terms in a terminological network to reduce its complexity. These clusters are then manually processed by a domain expert to either edit them or reject them.

Several distance metrics were proposed in the literature and can be applied to measure the quality of context extraction. Prior work had presented methods based on information retrieval techniques (Rijsbergen, 1979) for extracting contextual descriptions from data and evaluating the quality of the process. Motro and Rakov (1998) proposed a standard for specifying the quality of databases based on the concepts of soundness and completeness.

The method allowed the quality of answers to arbitrary queries to be calculated from overall quality specifications of the database. Another approach (Mena, Kashyap, Illarramendi, & Sheth, 2000) is based on estimating loss of information based on navigation of ontological terms. The measures for loss of information were based on metrics such as precision and recall on extensional information. These measures are used to select results having the desired quality of information.

We propose the use of a fully automatic context recognition algorithm that uses the Internet as a knowledge base and as a basis for clustering.

1.4 Outline

This paper constructs a model for context recognition using the Internet as a knowledge base. Section 2 presents a formal definition of context recognition and divides the problem into two subproblems. Section 3 builds the context recognition model, which consists of four main processes: the collection of data, the selection of possible contexts for each text, the ranking of the contexts, and the declaration of the current context, and describes the use of the Internet as a context database. Section 4 evaluates the model, using different types of input—Internet chats, e-mails, article abstracts, and Shakespeare plays. Section 5 discusses the results of the algorithm, possible applications, and directions for future research. Finally, Section 6 presents some conclusions.

2 Formal definition of the problem

After the review of the related literature on context, the following question can be addressed: How can context recognition be implemented based on using the Internet as a knowledge base?

McCarthy and Buvac (1997) formalized context as first class objects with the following basic relations:

ist(c,p) meaning that the proposition p is true in the context c, and
value(c,e) designating the value of the term e in context c.

The context objects were introduced as abstract mathematical entities with properties useful in artificial intelligence. However, in their paper the unique conclusion about what context is was avoided.

In this paper context is defined as any textual description that is most commonly selected by a set of subjects to describe a given situation:

Let $P_1,\ldots, P_m$ be a series of textual propositions defining situation S. Contexts $C_1,\ldots, C_k$ are defined as the context of situation S if $\exists$ n subjects, $n\geq 1$ so the majority of n selected ist(C_i, P_j) $\forall$i, for a given j (Contexts C_1 ,..., C_k are true for textual proposition P_j)

For example, assume there are three text propositions: "suspect", "body", and "knife". Assume also that two contexts exist, "murderer" and "not murderer". Propositions "suspect" and "body" are true in context "murderer". Propositions "suspect" and "knife" are true in context "not murderer". The two seemingly contradicting contexts can be true according to the existing data. This apparent contradiction is the situation in which a police detective or a judge is found, when the information is insufficient for an unequivocal decision and thus two possible directions can be pursued.

The definition of context allows the integration of statistical measures of context analysis using the terms of context formalization presented by McCarthy and Buvac.

For a series of propositions there exists a collection of sets of contexts.

Let $P_1, \ldots, P_m$ be a series of textual propositions when $\forall P_i$ there exists a collection of sets of contexts C_{ij} so that: For each i, ist(C_{ij}, P_i) $\forall$j meaning that the textual proposition P_i is true in each of the set of contexts C_{ij}. C_{ij} are not predefined hierarchically in a structure such as a tree. However, hierarchical structures can be built according to a specific set of textual propositions.

The main research problem is formally defined as:

What is the outer context C defined by

$$\mathrm{ist}\left(C, \bigcap_{i=1}^{m} \mathrm{ist}\left(C_{ij}, \mathrm{P_i}\right)\right)\forall \mathrm{j}$$

The number of existing contexts is assumed to be finite and to satisfy

$$C, C_{ij} \subseteq U_c \text{ (unity of all existing contexts)}$$

The main research problem can be divided into two subproblems:

1. Let P be a given single text. What are the possible contexts C_i that satisfy ist(C_i, P) $\forall i$ (for single text P all contexts C_i are true)

   Problem 1 assumes that some text is received. Each text can have a set of contexts that are true for the text. For example, a text saying "small step" can create the context of children, dwarfs, technological advances, and the Apollo moon landing. The problem is how to select these contexts from the unity of all existing contexts. This is the stage of the extraction of the possible contexts.
2. Let $P_1, \ldots, P_m$ be a set of texts that satisfy the following condition: for each text P_i there exists a set of contexts C_{ij} so that ist(C_{ij}, P_i) $\forall$j. What is the outer context C so that $\mathrm{ist}\left(\mathrm{C}, \bigcap_{i=1}^{\mathrm{m}} \mathrm{ist}\left(\mathrm{C_{ij}}, \mathrm{P_i}\right)\right)\forall \mathrm{j}$

Problem 2 assumes that you have a series of texts, each one with a series of contexts. Assume the above "small step" with the above list of contexts is given and another text "astronaut" is given with a list of contexts: space, NASA, moon landing, mission to mars, and want to be an astronaut. How can the possible number of contexts based on these two texts be minimized to "children who want to be an astronaut" and "moon landing"? Which of these two contexts is more likely to represent the two texts that are mentioned or is there a context that can include both contexts fully or partially?

This is the stage of the refinement and specialization of the set of possible contexts so that the best context or contexts can be chosen according to the rank that each context receives.

In the first example, there are three text propositions “suspect”, “body”, and “knife” and there are two existing contexts “murderer” and “not murderer”. Propositions “suspect” and “body” are true in context “murderer” while propositions “suspect” and “knife” are true in context “not murderer”. Thus, the outer context would include both “murderer” and “not murderer”. The two contradicting contexts can be explained as a result of a lack of information, when further information is required to determine which context receives higher ranking.

3 The context recognition model

The paper develops a model of context recognition. The research model is based on the streaming in text format of information that represents input from different sources—Internet chats, article abstracts, or Shakespeare plays. The information input to the context recognition algorithm can include any textual data that describes the situation such as speech, description of the actions performed by the speakers, facial expressions, background scenery, and nonactive participants in the situation.

The context recognition model output is a set of contexts that attempt to describe the current situation most accurately. The set of contexts is a list of words or phrases, each describing an aspect of the situation. The algorithm attempts to reach results similar to those achieved by the human process of determining the set of contexts that describe the current situation. The model, which is outlined in Fig. 1, consists of four major processes:

- Collecting Data—Each textual input item from the information sources is placed into a set of keywords.
- Selecting Contexts for Each Text (Keywords)—For each keyword a set of preliminary contexts is extracted from the Internet, which is used as a context database.
- Ranking the Contexts—Each preliminary context is ranked according to the number of references it receives in the context database and the number of appearances it has in the text.
- Declaring the Current Context—The preliminary contexts that have significantly higher numbers of references and higher numbers of appearances will be included in the current context.

The process of determining the current context continues as long as information about the current situation continues to stream in. The system constantly confirms and disconfirms the possible contexts as related to the different items of information received.

The algorithm determining the context was modeled and implemented in Java code. The context recognition model input was evaluated using different fields of knowledge: computer Internet chats, health Internet chats, e-mail messages to a local government, physics research abstracts, and Shakespeare plays.

3.1 Collecting data

The Blackboard architecture is implemented in the context recognition model. Each new textual data item of the current “world state”, the situation, serves as input. All the texts describing the current situation are treated similarly. The data received appear as a set of words that are organized as single words or sentences.

The text was not preprocessed. All misspelled words were left in the text. The text format determines the amount of input for each step or loop of the algorithm; a long set of words is treated the same as a single word sentence. Each step processes the words

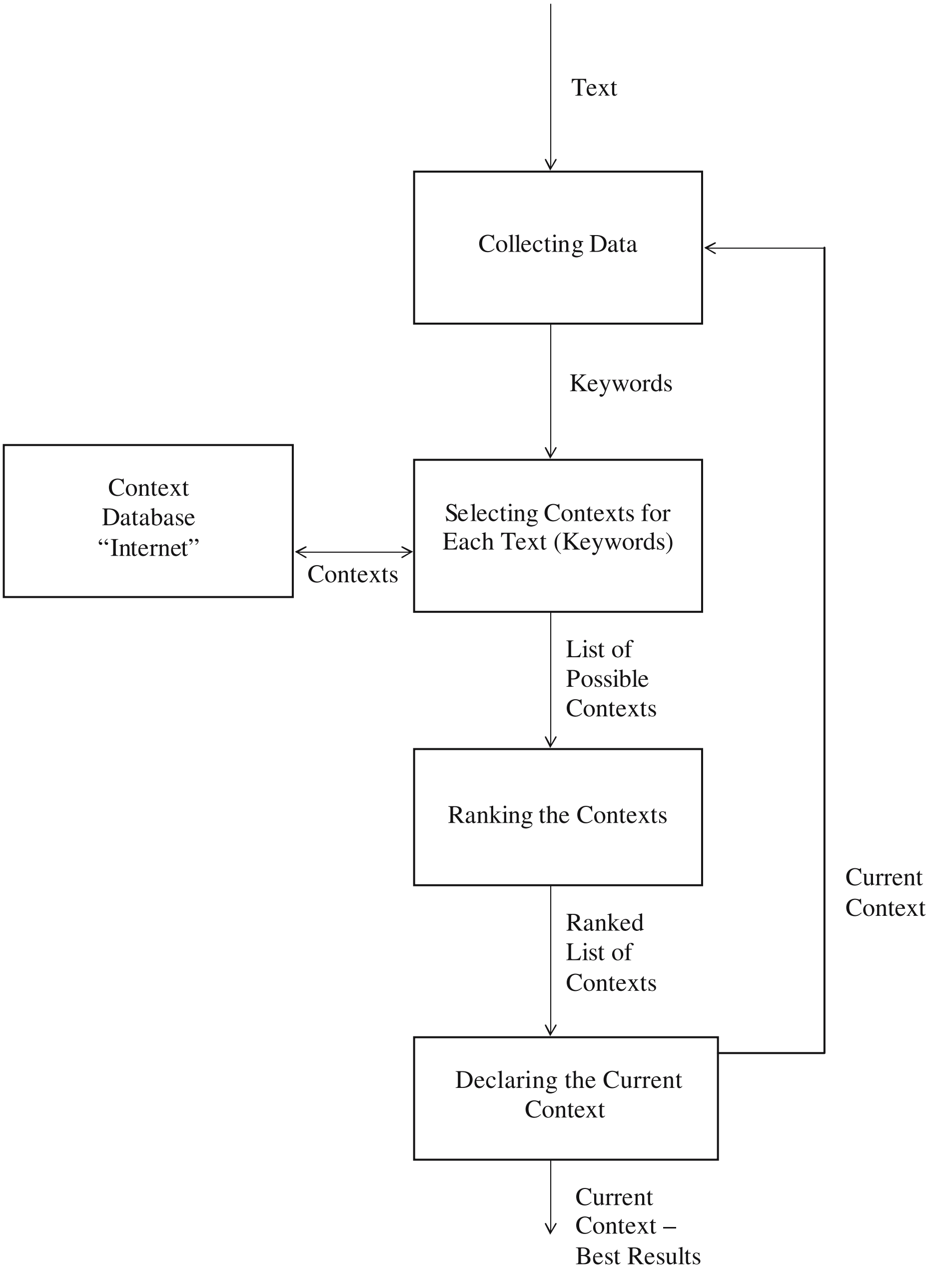

Fig. 1 The context recognition model outline

received between new line characters or punctuation marks. Consequently, the input of each step can consist of a sentence, a part of a sentence, or a change in speech in a conversation. However, in cases of long sentences without any punctuation marks, the processed data was sliced according to the maximum number of word inputs that could be processed by the search engines used in the following steps.

Each text is decomposed into single words. Words are letter strings separated by spaces. All punctuation is removed from the text. The words are then checked through a set of dictionaries. The first dictionary can be referred to as the "Stop List". This dictionary includes all the words that do not add to the understanding of the context, such as: I, me, in, are, the. All words that appear in this dictionary are ignored. The next step includes a set of dictionaries according to fields of knowledge. The dictionary is used to sieve the words that are not related to the specific field of knowledge. The application of the algorithm uses three different types of dictionaries for the evaluation. However, any dictionary can be used or built for this purpose. The dictionary used for the computer related chats is the Foldoc Dictionary of Computing (http://foldoc.doc.ic.ac.uk), for the health related chats the On-Line Medical dictionary (http://cancerweb.ncl.ac.uk/omd), and for the physics abstracts the World of Science Wolfram web resource (http://scienceworld.wolfram.com). If the word appears in the field of knowledge dictionary, then it is added to the list of keywords that are searched in the context database. If the word does not appear in the field of knowledge dictionary, then it is ignored.

The evaluation also examined the implementation of the algorithm without the use of any knowledge dictionary. Therefore, for the Shakespeare plays no dictionary is used to check the performance of the algorithm when the field of knowledge is not known.

This process continues for each word in the text. After each text passes through this module of the algorithm, a list of words is sent to be checked for a possible set of contexts.

3.2 Context database

The Internet was used as a context database for the selection of the context of the input text. The contexts were represented by words or sets of words. The words or sets of words representing the contexts can be viewed as meta data created for each set of Internet web pages. These words can be created online using existing clustering methods or can be predefined by search engines.

The Internet can then be viewed as an immense set of words that represent different possible contexts, each associated with its respective web page. For each query checked on the Internet using a search engine that looks for keywords, a list of preliminary contexts is obtained related to each web page that is retrieved. The preliminary contexts are created by applying a term frequency method on the web pages retrieved. Each web page is checked for frequent terms that represent possible contexts. These terms are the preliminary contexts.

Internet technology can also be used to implement a database directed only at a limited field of knowledge. For example, an Intranet can be built containing information on computer technology. These web-based files can serve as a context database for situations related to computer technology. Similarly, a newspaper archive can serve as a context database for classifying new articles.

3.3 Selecting contexts for each text

The selection of the current context is based on a search through the database for all relevant documents according to keywords and on the clustering of the results into possible contexts. If the Internet is used as a context database, then any existing search engine can be integrated into the algorithm to yield a set of documents. Each document is associated with a set of preliminary contexts that can be matched with it.

The method used for clustering the contexts is based on Term Frequency/Inverse Document Frequency (Salton, 1989). The method is a term-based approach that focuses on the relationship of the documents to the corpus. This method checks the frequency of words appearing in the text and the absence of words appearing in the text. This method was selected for its ease in retracing the context to the original text and its ease of implementation.

The application of the algorithm is based on the concise all pairs profiling (CAPP) clustering method (Valdes-Perez & Pereira, 2000). This method approximates profiling of large classifications. It compares all classes pairwise and then minimizes the total number of features required to guarantee that each pair of classes is contrasted by at least one feature. Then each class profile is assigned its own minimized list of features, characterized by how these features differentiate the class from the others.

The CAPP clustering method is implemented in Vivisimo (http://www.vivisimo.com), a search engine that returns search query results in clusters by keeping track of the names of the groups while forming them. If a cluster forms during the clustering stage and cannot be described, then the cluster is rejected.

However, many other search engines cluster their search results and can be integrated in the algorithm. Other search engines can be implemented to select the context using their hierarchical tree format of storing web files as sets of contexts. The advantage of the Vivisimo search engine results, which are based on the URL, description, and document titles, is their easy accessibility.

Once a list of keywords exists, each keyword is searched in the context database—the Internet. This creates a list of preliminary contexts for each keyword. The full list of preliminary contexts for all the keywords includes all the possible contexts for this current text.

An example of this process is presented in Section 3.6.

3.4 Ranking the contexts

The ranking algorithm is described in Fig. 2. The algorithm checks the number of appearances in the text for each preliminary context in the Set of Preliminary Contexts. The contexts are also examined for the number of Internet documents that refer to the Set of Documents. The set of contexts is now ranked according to both the number of references in the text and the number of references in the documents (Fig. 3).

These two metrics were selected since the number of appearances in the text represents how many times each preliminary context was mentioned in the situation. The number of references in the Internet represents how important the preliminary context is to the general population that uses the Internet.

New preliminary contexts can now be created according to textual substrings of existing preliminary contexts. They can also be united according to singular and plural form or according to similarities in meaning—synonyms. The contexts are united by adding the number of appearances and the number of references: the ranking of the plural form is added to the singular form and, in the case of similar meanings, the ranking of both contexts can be used for each one.

The text data input is streamed, creating a continual process that extends the list of possible preliminary contexts. Finally, the information also flows back, allowing the current context to be constantly updated.

This step in the algorithm has three inputs: the set of preliminary contexts received from the previous stage as an unranked list, the text that includes all the information fed into the algorithm so far, and the set of documents that refer to the set of preliminary contexts.

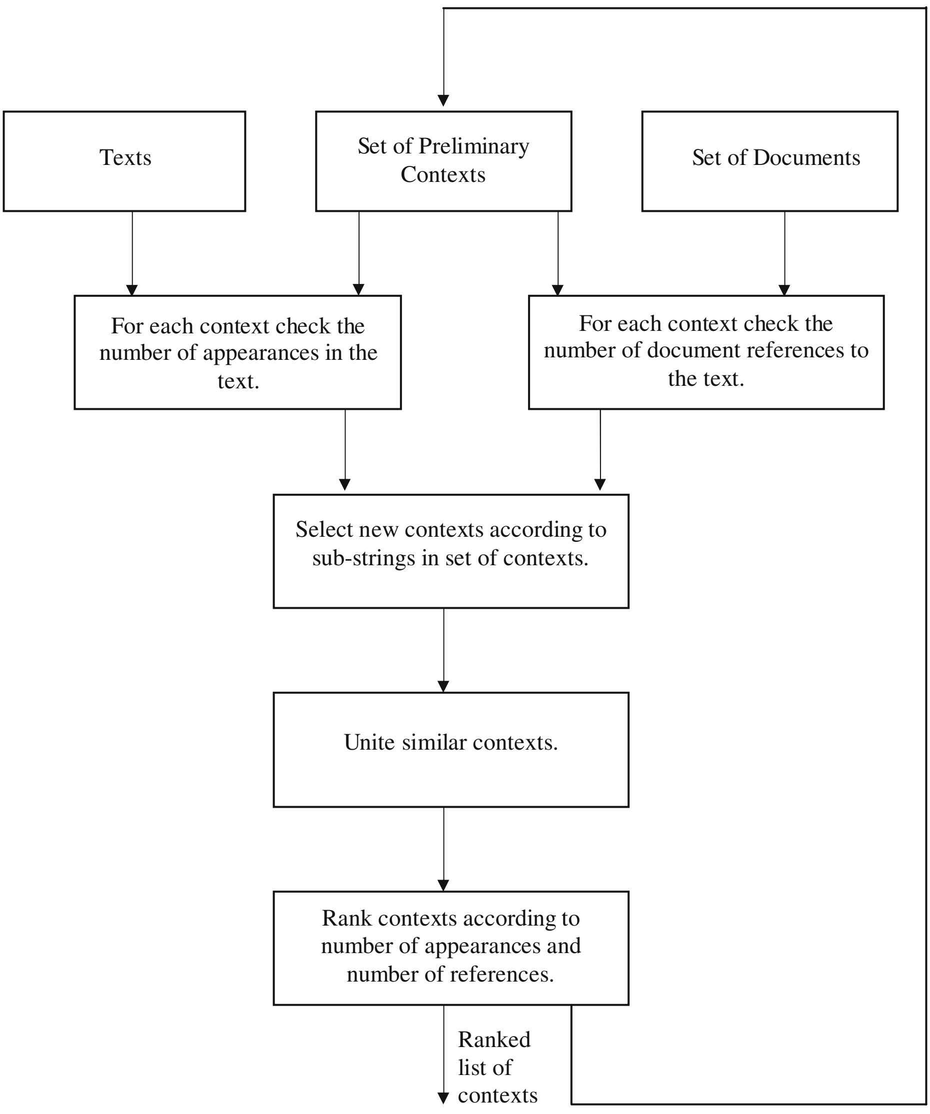

Fig. 2 Ranking algorithm

The first step sums the number of times each preliminary context was referred to in the text. Each time a preliminary context is mentioned in relation to the text it receives a higher ranking in the number of appearances.

Similarly this step sums the number of documents referring to the preliminary contexts. Multiple reference pages from similar web sources are counted as one instance. Each document usually refers to multiple contexts, consequently creating a long list of preliminary contexts.

New preliminary contexts are selected according to substrings of existing preliminary contexts. Substrings are any part of a context that can be broken down into single words. To

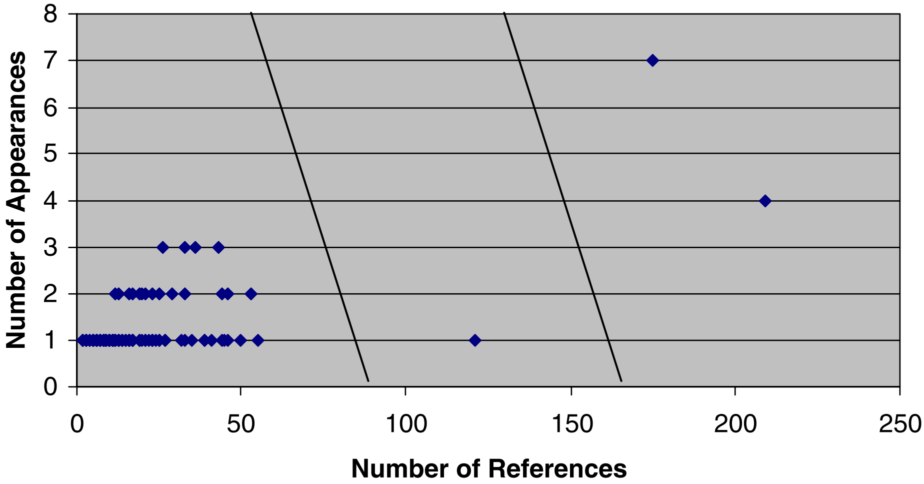

Fig. 3 Declaring the current context

limit the number of short substrings that were parts of the original string but have no semantic relation to it, a minimum number of characters per substring can be defined.

However, the substrings should be filtered to prevent erroneous contexts by using "Stop-List" techniques. These new preliminary contexts receive the sum of the ranking values of all the preliminary contexts that contain the substring.

The next step includes uniting similar preliminary contexts. The algorithm unites preliminary contexts according to singular and plural forms of the contexts. Another way of uniting is according to similarities in meaning, synonyms, using a thesaurus.

The last step involves ranking the set of preliminary contexts according to both the number of references from the documents and the number of appearances in the text. This step maps all the preliminary contexts to a two dimensional graph, allowing the contexts that receive very high ranking in both characterizations to be located.

After each session of ranking, the list is used for two purposes—readjusting the ranking values of the set of preliminary contexts and declaring the current context. The current list forming the context joins the new preliminary contexts arriving from the continuously streaming text. The lists are united and the ranking process is repeated. In parallel to the repetition of the ranking algorithm, the set of ranked preliminary contexts is forwarded to the next module to determine the current context.

3.5 Declaring the current context

The output of the ranking stage is the current context or a set of highest ranking contexts that differ essentially. The algorithm then returns to the first step to collect more texts and feed them again to the database. The set of preliminary contexts that has the top number of references, both in number of Internet pages and in number of appearances in all of the texts, is defined as the highest ranking and is declared to be the current context.

The current context received from the previous stage can be depicted on a graph according to number of appearances and number of references as in Fig. 3.

The algorithm for detecting the current context includes the following steps:

1. Organize the list of preliminary contexts in descending order according to number of references appearing in the Internet—the Set of Documents.
2. Find the difference between each value of the number of references and its nearest lower value neighbor.
3. Find the difference between each value of the number of appearances and its nearest lower value neighbor.
4. Weight the number of appearances in the text and the number of references in the Internet according to the following formula:

MVR	Maximum value of references
MVA	Maximum value of appearances
CRDV	Current reference difference value (calculated in step 2)
CADV	Current appearance difference value (calculated in step 3)

$$\text{WeightedValue} = \sqrt{\left(\frac{2^{*}CADV^{*}MVR}{3^{*}MVA}\right)^{2} + (CRDV)^{2}}$$

   The index of the number of references is on a much larger scale than the index of the number of appearances, and therefore it is not possible to retain the original proportions and it is necessary to rescale the indices. The Weighted Value can be viewed as the weighted distance to the origin. A constant of two thirds was found to be appropriate for the readjustment of the figures.
5. Find the maximum value of the Weighted Value. If the maximum Weighted Value is the first value, then continue to the next one, since frequently the first value is too far from its neighbor.
6. Select all the contexts that appear before the maximum Weighted Value in the list that was organized in step 1 as the current context. Store current selected context.
7. Erase the selected contexts from the list and repeat steps 5 and 6 again.

The first cluster of contexts near the origin includes all the contexts that received low ranking both in number of appearances and in number of references. This group of contexts includes most of the contexts in the list. Since the contexts in this group received low ranking they are eliminated from the list. The remaining contexts are the current context.

The process can continue until all the contexts in the list are covered and this will yield all the possible preliminary contexts. However, in most cases the best results were already achieved when steps 5 and 6 were performed two times. Further repetitions, which increase the number of results, were unnecessary.

In this paper, the weighted value was used for ranking, since it yields the best ranking results. However, if the ranking is performed according to the number of references and not according to the weighted value, the results are better than when ranking according to the number of appearances. This indicates that the number of appearances of the context in the text has less value in the determination of the context than the number of references in the Internet. We assume that this is the case, since when using the Internet as a Context Database, the vast size of that 'database' allows for a larger sample set of documents, which results in a more accurate sense of contextual meaning.

During the implementation of the algorithm, there was a problem that required special consideration. The contexts that received lower ranking than the top ranking contexts in the cluster but were not part of the cluster were kept. Namely, these are contexts that receive lower ranking in either the number of appearances or the number of references than the top ranking contexts, but not in both. Running the algorithm showed that these contexts are sometimes relevant and should be kept.

The current context is the output of the algorithm. However, since the algorithm is continuous, the contexts continue updating as long as new textual input continues to be accessed by the algorithm.

3.6 Example of algorithm implementation

3.6.1 Example of health-related chat context evaluation

Jane: NO I DONT THINK SO IVE DONE ALOT OF SEARCHING THE NET FOR INFO
Jane: HAVE GATHERED ALOT
Lisa: are you a diabetic pt.?
Jane: YES
Jane: TYPE 2
Jane: 1 YR
Jane: AND A BIT OVERWEIGHT BUT NOT OBESE
Lisa: ok. are you married?
Jane: YEP
Jane: 23 YRS
Lisa: do you have children?
Jane: HAPPILY MARRIED
Lisa: do u control your blood glucose level
Jane: YES DIET AND MEDICINE
Jane: AND EXERCISE
Lisa: is there any complications like eye problems
Jane: I WEAR GLASSES

The input is read one line at a time. Each word is separated by a space. Punctuation marks are eliminated. Each word is checked against the "Stop List" dictionary. In this case each word was checked in a predefined medical dictionary.

The words that passed the previous stage serve as keywords. After each step (change of speaker), the keywords are sent to the search engine and clustered into a list of preliminary contexts.

These steps are repeated 17 times, yielding 139 preliminary contexts that have at least two references in the Internet and are relevant to keywords that appeared at least once in the text. These preliminary contexts can also be words that did not appear in the text itself but were a result of the clustering of the web pages.

The contexts are clustered according to substrings, plural form, and similarities. For example:

> American Diabetes Association (7,1) and Diabetes Type 2 (121,1) are clustered with Diabetes (81,2) forming the ranking of (209,4) for Diabetes. Similarly, the contexts of Eye Care and Common Eye Problems are clustered with Eye.

The algorithm yielded the following ranked contexts results for the chat.

The current context includes Diabetes (209,4), Eye (175,7), and Diabetes Type 2 (121,1). The values in the parentheses are the number of references and the number of appearances respectively.

The following graph shows the weighted value calculated for each context (Figs. 3 and 4). The differences show that the maximum weighted value is after Eye (69.66), resulting in two contexts: Diabetes and Eye. The second maximum weighted value is after Diabetes Type 2 (67.61), resulting in only one additional context, Diabetes Type 2. As a result the algorithm selected the first three contexts—Diabetes, Eye, and Diabetes Type 2—as the current context.

The remaining contexts form a new graph, which can again be clustered. This process can be continued, yielding additional contexts. However, usually two repetitions are sufficient to obtain the context.

4 Context model evaluation

The context model algorithm was evaluated using different fields of knowledge: computer Internet chats, health Internet chats, e-mail messages to a local government, physics research abstracts, and Shakespeare plays.

The field of computer Internet chats was the field most extensively analyzed. These chats were acquired from the MSN chats. The chats included a few participants and were observed over time. Parts of the chats that dealt with topics concerning computers were copied to files. From these chats sets of files were randomly selected to be analyzed by the algorithm. These files were fed as input to the context algorithm. The results were compared with the results given by computer literate subjects.

The subjects who answered the survey were graduate students with at least basic knowledge of computer terminology. The students were asked whether they have prior knowledge in programming in at least one programming language. Some of the students who participated in the survey were attending an Artificial Intelligence course. Other students who participated in the survey were selected from the computer lab and were asked whether they had computer background and knew at least one programming language before they were requested to take part in the survey.

A total of 20 subjects participated in the survey. The participants received a set of three chats. This allowed two groups of participants to be formed, each group with a different set

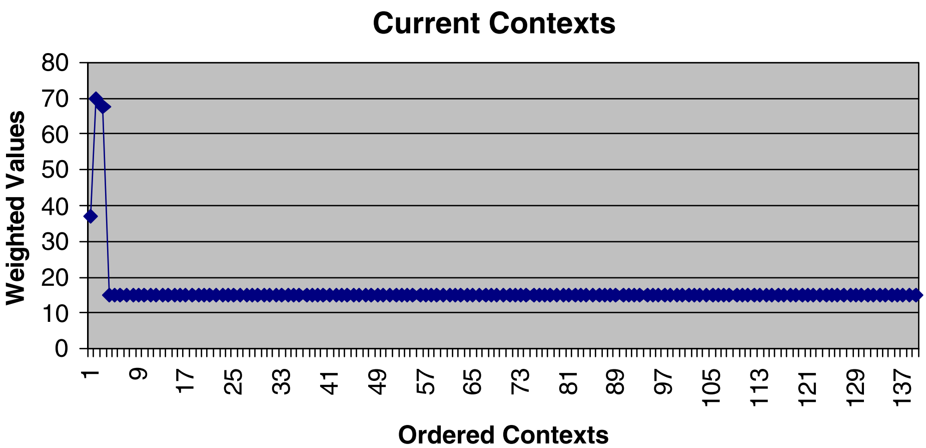

Fig. 4 Weighted values of the current contexts

of chats. The division into two groups of participants was performed to extend the set of sample data examined and monitor the performance of the groups. Each chat had at least nine replies. The maximum number of replies per chat was 11. An average of ten subjects determined the list of words forming the context for each chat. A total of six chats in computer related topics were analyzed in this way.

The subjects were presented with the above chats and were asked to provide a list of words describing the contexts for the chats. The subjects were told that the text was obtained from Internet chats and was presented to them as is, including spelling mistakes and Internet acronyms. The subjects were asked to write in their own words what they felt were the context.

Contexts were defined to the subjects as any word or set of words that come to mind when reading the text that best describes the situation presented. The number of words was not limited in length. The words describing the context were not selected from a list. The subjects could write any word that came to mind when they described the text presented to them.

The words were counted to determine the number of times they appeared among all the subjects. Each word was counted once for each subject who mentioned the context. The context was ranked in descending order according to the number of times that the subjects mentioned the context.

The list of best-ranked contexts was compared with the results yielded by the context recognition algorithm. However, the other contexts that were mentioned more than once by the subjects were also compared to check the sensitivity of the algorithm. The comparison of the performance between the two groups of participants showed similar results.

The randomly selected health chats were presented to two medical doctors, who were asked to define the health topic discussed. Similarly to the previous experiment the doctors were asked to select the context as a word or set of words best defining the situation defined in text. They determined the contexts of the chats and their results were compared to the results obtained by the context recognition algorithm.

The context recognition of e-mail messages written to and by a local government was also evaluated. These e-mails, in German, addressed the Perspectives du Theatre Festival event and consisted primarily of e-mails from citizens to the city hall or press releases and announcements from the city outward. The challenge was to analyze this material and provide a useful set of classifications so that the materials could be rapidly understood and sent to the appropriate people for response.

The goal of the topic classification experiment was to identify the topic of the e-mail according to a predefined list of topics supplied by Saarbrücken that relate to the Perspectives du Theatre Festival. The predefined topics of the e-mails supplied by Saarbrücken were: Organisation, Veranstalter, Finanzen, Besucher, Informationen, Rahmenprogramm, Spielplan, Other. Each topic was accompanied by a set of words, contexts, which describe it. All the data was supplied by Saarbrücken in the German language.

The evaluation of context recognition in the field of physics differed from the evaluation process in the previous two fields of knowledge and included running the algorithm on research abstracts in this field. The research abstracts were taken from a summary of articles sent by e-mail to a mailing list of physicists. After the algorithm was run on the research abstracts, two physicists from the mailing list were asked to rank both the results and the overall performance of the algorithm. The performance of the algorithm was defined as successfully identifying the context, as a set of words, defining the research abstracts.

Famous scenes from Shakespeare plays were selected and used to check how the algorithm detects the context without any information in the field of knowledge with which it is presented. Two whole scenes taken directly from Shakespeare's writings were input for the algorithm.

Table 1 summarizes the algorithm evaluation process and presents the field of contexts, the purpose of each evaluation process, the number of cases evaluated in each field, the number of subjects evaluating the results, how the subjects were selected, and the results.

Tables 2, 3, 4 and 5 addresses the results in the field of the computer chats. Table 2 summarizes the results received by the algorithm and the results received by the subjects. The table also displays the average results of the context recognition model for all the six cases examined. The second and third sets of contexts mentioned by the subjects are also compared with the second and third reiterations of the algorithm to check the sensitivity of the algorithm.

The top-ranking contexts mentioned by the subjects were identified as contexts by the algorithm. *In most of the cases the contexts that were ranked among the highest by the subjects were also ranked among the highest by the algorithm.*

Some of the other contexts generated by the algorithm were not selected by the subjects. In addition, a few of the lower ranking contexts mentioned by the subjects were missed by the algorithm.

Table 2 shows that for the top ranking contexts the algorithm yields very high results. As more contexts that received lower ranking by the subjects are added, the results of the algorithm degenerate. The algorithm needs improvement in deducting better results in the second and third ranking sets of contexts.

Similar tests were performed for subgroups of n−1, n−2, n−3, and n−4 subjects. For example, if there were ten subjects in the group, all possibilities for 9, 8, 7, and 6 were examined. The results were averaged for all of the subgroups. The mean, standard deviation, variance, and number of groups for each of the cases and for each ranking level are displayed in Table 3.

Table 1 Summary of evaluation process results

Context field	Purpose	Number of cases	Number of subjects	Subject selection	Results
Computers	Statistical analysis	6	20 (10 for each case)	Computer literate	Algorithm results are similar to human subjects determining the context
Health	Check if algorithm can work in other fields of knowledge	2	2	Doctors	The algorithm can be used in the field of medicine
Physics	Check if algorithm can work in other fields of knowledge	3	2	Physicists	The algorithm can be used in the field of physics
Shakespeare Plays	Check how the algorithm can detect context without any information in the field of knowledge	2	–	–	The algorithm can be used without any information in the field of knowledge
E-mail messages to/from local government	F score comparison with other techniques	104	3	Municipality workers	The algorithm achieves high F score compared to other techniques

Table 2 Ranking of the context recognition algorithm

Context recognition	Ranking						
	Chat 1	Chat 2	Chat 3	Chat 4	Chat 5	Chat 6	Average
Top ranking contexts	100%	100%	100%	100%	100%	100%	100%
Top and second ranking contexts	100%	84.46%	65.22%	57.15%	75%	100%	80.31%
Top, second, and third ranking contexts	86.67%	81.25%	57.15%	40%	76.92%	75%	69.50%

The summarized results displayed in the table above show that, even when the size of the sample is extended through the examination of additional possible subgroups, the results yield a high percentage of correlation between the contexts identified by the group and the contexts yielded by the algorithm.

The significance of the results was analyzed using the identical populations test. The test for homogeneity is designed to test the null hypothesis that two or more random samples are drawn from the same population or from different populations, according to some criterion of classification applied to the samples.

The Chi-square Pearson Test for Association is a test of statistical significance. Typically, the hypothesis is whether or not two populations are different in some characteristic or aspect of their behavior based on two random samples.

The algorithm replaced one of the subjects in the group. The results of the new group of subjects and the algorithm were analyzed to determine the new set of contexts and their rankings. The process was repeated for each of the subjects replaced by the algorithm. The average result of all these groups was compared with the original group that contained only human subjects. Table 4 displays the identical population test results for each of the six chats.

Table 3 Ranking of the context recognition algorithm—subgroups results

Context recognition	Ranking						
	Chat 1	Chat 2	Chat 3	Chat 4	Chat 5	Chat 6	Average
Mean							
Top ranking contexts	100%	97.39%	80.21%	99.74%	96.07%	100%	95.57%
Top and second ranking contexts	100%	87.92%	67.69%	65.65%	83.44%	100%	84.12%
Top, second, and third ranking contexts	91.42%	82.44%	58.42%	44.98%	78.59%	80.87%	72.79%
Standard deviation							
Top ranking contexts	0.0	8.54	36.51	2.55	12.80	0.0	–
Top and second ranking contexts	0.0	11.05	21.24	32.75	15.30	0.0	–
Top, second, and third ranking contexts	6.5	6.90	8.79	15.02	8.56	16.11	–
Variance							
Top ranking contexts	0.0	0.73	13.33	0.07	1.64	0.0	–
Top and second ranking contexts	0.0	1.22	4.51	10.73	2.34	0.0	–
Top, second, and third ranking contexts	0.42	0.48	0.77	2.26	0.73	2.6	–
Group count							
Top ranking contexts	386	386	352	96	386	155	–
Top and second ranking contexts	386	386	562	549	386	214	_
Top, second, and third ranking contexts	386	386	562	562	386	256	–

Table 4 Identical populations test

H0: The populations are almost identical. There is no significant difference between the populations			
Chat 1	χ^2=0.065199	*P* value=1	Little or no real evidence against the null hypothesis
Chat 2	χ^2=0.568693	*P* value=0.999	Little or no real evidence against the null hypothesis
Chat 3	χ^2=0.187391	*P* value=1	Little or no real evidence against the null hypothesis
Chat 4	χ^2=0.256795	*P* value=1	Little or no real evidence against the null hypothesis
Chat 5	χ^2=0.133712	*P* value=1	Little or no real evidence against the null hypothesis
Chat 6	χ^2=0.273300	*P* value=0.998	Little or no real evidence against the null hypothesis

The same test for identical populations was repeated for all subgroups containing n−1 subjects. For example, if there were ten subjects in the original group, then now all possibilities of nine human subjects and the algorithm replacing each of the nine subjects were analyzed. The average result of all the possibilities was compared to the original group consisting of only human subjects, displayed in Table 5.

The results of the identical populations test comparing the groups containing the algorithm as a subject with the original group consisting only of human subjects showed that they were almost identical populations. In other words, if the computer is part of the group, the context will remain identical. **Hence, there is no significant difference in the determination of context between the algorithm and the human subjects.**

The examination of the results of all the subgroups of n − 1 subjects yields a similar conclusion. All results showed that the populations were almost identical. Thus, even in the examination of the smaller populations in which the algorithm has greater impact on the determination of the contexts, the results remained that the populations are almost identical. **Therefore, the examination of the subgroups in the determination of context showed that if the algorithm replaces one of the subjects there is no significant difference between the populations.**

Table 6 compares the results received by the algorithm and the results received by the subjects. Context is defined as a text description of a situation most commonly selected (best ranked) by a set of subjects. In this example, as a result of ten participants analyzing the chat, a context is defined as when six or more participants selected the textual description. However, the other contexts that were mentioned more than once by the subjects are also compared to check the sensitivity of the algorithm.

Table 6 displays the ranking of the context recognition based on the rank of number of the successful contexts identified by the algorithm out of the total contexts mentioned by the subjects. The context receives ranking of 100% if all of the human generated results are covered by the system generated results. The number of textual descriptions selected by the algorithm usually reaches up to ten descriptors. In extreme cases where the context of the

Table 5 Identical populations test for subgroups

H0: The populations are almost identical. There is no significant difference between the populations			
Chat 1	χ^2=0.483390	*P* value=0.993	Little or no real evidence against the null hypothesis
Chat 2	χ^2=3.981739	*P* value=0.782	Little or no real evidence against the null hypothesis
Chat 3	χ^2=1.507506	*P* value=0.997	Little or no real evidence against the null hypothesis
Chat 4	χ^2=2.274965	*P* value=1	Little or no real evidence against the null hypothesis
Chat 5	χ^2=1.183560	*P* value=0.978	Little or no real evidence against the null hypothesis
Chat 6	χ^2=1.829639	*P* value=0.872	Little or no real evidence against the null hypothesis

Table 6 Comparison of the results of the algorithm and the human subjects for chat 1

Algorithm (computer algorithm)	Subjects (human respondents)	Rank (number of respondents)	Match
Java	Java	10	Yes
Help/FAQ	Help	6	Yes
Program	Program	5	Yes
Source code/Software	Code	5	Yes
Design	Programming	2	No
Buy	Problem	2	No

situation described is less clear and more vague the context can reach 20 descriptors or more.

The following tables show additional information related to the performance of the algorithm in other fields of knowledge. Table 7 displays the overall ranking of the algorithm in identifying Physics abstracts. Table 8 displays the performance in identifying health related chats. Table 9 presents the ability of the algorithm to perform without any prior information of the field of knowledge using Shakespeare plays.

In comparison to other simple key phrase extraction models based on the number of words appearing in the text, the algorithm performs much better. In the example of the health related chat, counting the number of word appearances would result in the words "And" and "You" as possible contexts. Diabetes and Eye would not be selected since they appear only once. Diabetes Type 2 would not be selected since it is based on a set of different words that are separated by different speakers.

In comparison with more complex methods such as Dumais and Chen (2000), who use a hierarchical structure for classifying a large, heterogeneous collection of web content, the algorithm performed much better. Their overall *F* Score value for the top-level categories was 57.2%. They mentioned that the performance on the original training set was 64.9%.

An experiment conducted on data of the local government of Saarbrücken analyzed the materials by topic (ticket/travel information, finances, organization, etc.). The systems' performance achieved high correspondence to human results for the different topics. The experiment included 104 different e-mails to analyze context recognition. Table 10 summarizes the comparison of the results of the context recognition algorithm to human judgments.

5 Discussion

Current information seeking and information retrieval applications found on the Internet use clustering techniques to present their results. The present research automates context recognition. This application of context recognition does not classify material found on the

Table 7 Physics ranking of the overall performance of the context recognition algorithm

Context recognition	Ranking			
	Abstract 1	Abstract 2	Abstract 3	Average
Context recognition ranking (out of 7)	5	4.5	5	4.83

Table 8 Algorithm performance in health related chats

Context recognition	Ranking	
	Chat 1	Chat 2
Top ranking contexts	100%	100%

Internet, but uses the Internet to obtain current context for the application using the algorithm.

The current search engines basically perform the opposite task of the context recognition algorithm. A user who uses a search engine provides a word or a set of words that is expected to specify the context of information he is interested in. The search engines uses the word or set of words to extract as many documents as possible that match this predefined category. However, for the context recognition algorithm the input can be viewed as a textual document where the output is the keywords that the user is looking for. The algorithm extracts the topics of a textual document according to words that might not appear in the document itself.

Furthermore, even if the Internet search engine can receive a whole document as input and provide all the words representing meta tags or predefined topics of the Internet pages, the number of possible words retrieved would in most cases outnumber the words in the original document. The result would be an input document of up to a hundred words and an output of several hundreds of words describing the document, most of which are not relevant. For a person or a machine who are using this output it would be easier just to scan the original input.

As seen in the example in the previous chapter, if we use a simple key phrase extraction model, such as a search engine, we cannot retrieve the correct context of the document by simple term frequency. To achieve successful results a database is required. In models that use a predefined database, the results achieved were limited and much lower than those of the context recognition algorithm, which is based on the use of the Internet as a knowledge database.

The context recognition algorithm with the Internet as a knowledge base can be applied as a classification method for short movies or newspaper articles in a specific field of knowledge, such as computer science. The same algorithm can also be applied when a specified existing context database, other than the Internet, is given for the search for contexts. The implementation of the context recognition algorithm in either case would result in a context that most likely describes the situation.

The algorithm can also be used to classify for surveillance purposes chats held on the Internet. Each monitored conversation can be classified according to a set of relevant predefined contexts. The chats can be ranked according to their relevance to each context, thus allowing each chat to be investigated according to priority.

Table 9 Algorithm performance without information on field of knowledge

Context recognition	Ranking	
	Shakespeare play 1	Shakespeare play 2
Play name	Identified	Identified
Main players	Identified	Identified
Writer	Identified	Identified
Chapter in play	Identified	Not identified

Table 10 Comparison of the results of the algorithm for local government e-mails

	Context recognition
Precision	85.37%
Recall	84.34%
F score	84.85%

The complexity of the algorithm is *o(an)* where n represents the number of input cycles such as each line of text or each time that input is received from a different source. The "a" represents a constant limiting the number of top ranking results from each cycle of the algorithm. For example, "a" can represent top 10, 100, or 1000 best ranking contexts in each cycle. This allows different levels for the monitoring of the amount of data the algorithm handles.

The dynamics of the Internet (content) may cause different results for the same keyword. It is quite possible that there will be different results for the same keyword at different points in time. The research did not check the change of context as a result of change in time. It is possible that new contexts could be created as a response to topics that receive more public attention. The Internet is constantly changing. On the one hand, two different time periods can yield a change in the results. On the other hand, the change is a result of the importance of certain contexts related to world issues. The change of the number of pages relevant to a specific context can show a greater importance of this context even if it only appears once in the textual input. Thus, the Internet can be viewed as a constantly updating database.

The research can be extended to include a larger data set and data generated by other sources. Furthermore, the research can be expanded to include the integration into the context algorithm of existing systems, such as voice recognition or image identification systems. The integrated system could be tested with human subjects to see whether it recognizes the context of a situation that it encounters with its sensors.

Another possible field of research is automatic commonsense knowledge acquisition using context. Context recognition is a part of commonsense knowledge, which is one of the main problems in artificial intelligence. It may be possible to expand the technique used to build the context so that it also includes commonsense detection.

Natural language processing is based on the context of its input. Natural language algorithms and systems can be implemented using the context recognition algorithm. Such a system might be able to understand language more easily based on the known information on the current context.

The context recognition model, aside from its use of the Internet as a context database, is characterized by other features. First, the model works in real time with no training or practice required beforehand. Thus, it extracts the context immediately with little previous user intervention. Second, the model is marked by flexibility. It can function in multilingual environments, as noted in the paper with the presentation of implementation in the German language.

6 Concluding remarks

The development of artificial intelligence applications requires the careful consideration of the context. The research community has recognized this and has defined the problem of context recognition as one of the main questions. The main motivation of the paper is to present a model that allows applications to determine their context "on the fly", in real time, when today the context is predetermined as part of the predefined knowledge of the application. An application based on the model will be able to differentiate between its context information and the knowledge and data needed to perform its required tasks. The context recognition model will enable computers to communicate with humans more easily by providing the context of different situations.

The main idea of the research model was to use the Internet as a context database. The Internet is a source of information that is constantly increasing and being updated. The use of the Internet as a database for context recognition therefore gives a context recognition model immediate access to a nearly infinite amount of data in a multiplicity of fields. Hence, the necessity of creating a database for the determination of the context is eliminated.

Another main contribution is the iterative application of the clustering and the ranking algorithm, which allows new streaming text to be added and new contexts to be constantly developed. This more closely models the reality of searching, especially on the Internet, and may in fact allow for a learning mechanism at some later stage of research.

Furthermore, the situations for which the context is sought can be independent of the Internet; the Internet is merely the database in which the algorithm searches for the context. Thus, for example, the context of a conversation between people can be found through the use of the Internet—the algorithm is a tool that allows the computer to determine the context by using the Internet as a database and then to pass this context back into the real world.

The Internet is one possible source of data, but the algorithm holds also for a more restricted database. Intranet data, internally generated textual information about the organization that is stored, can also be used. Further research is necessary to determine the amount of data needed to implement the algorithm on Intranet data and to analyze the performance of the algorithm using restricted domain data. The algorithm achieves good context recognition results, both with and without the use of a field of knowledge dictionary, which represents specialized knowledge.

A model that imitates human context recognition will promote the understanding of human intelligence. This model will allow artificial intelligence to more exactly imitate the functioning of context recognition mechanisms.

References

AAAI (1999). *Workshop on reasoning in context for AI applications* (Workshop Series Tech. Rep. No. WS-99-14). Menlo Park: AAAI.

Aitchison, J., Gilchrist, A., & Bawden, D. (1997). *Thesaurus construction and use: A practical manual* (3rd ed.). London: Aslib.

Arens, Y., Knoblock, C. A., & Shen, W. (1996). Query reformulation for dynamic information integration. In G. Wiederhold (Ed.), *Intelligent integration of information* (pp. 11–42). Boston: Kluwer.

Assadi, H. (1998). Construction of a regional ontology from text and its use within a documentary system. *Proceedings of the International Conference on Formal Ontology and Information Systems* (*FOIS-98*). Amsterdam: IOS.

Buvac, S. (1996). Resolving lexical ambiguity using a formal theory of context, semantic ambiguity and underspecification. *CLSI lecture notes* (pp. 1–24).

Carver, N., & Lesser, V. (1992). Blackboard systems for knowledge-based signal understanding. In A. Oppenheim & H. Nawab (Eds.), *Symbolic and knowledge-based signal processing* (pp. 205–250). Englewood Cliffs: Prentice-Hall.

Dumais, S., & Chen, H. (2000). Hierarchical classification of web content. *Proceedings of SIGIR, 23rd ACM International Conference on Research and Development in Information Retrieval, Athens* (pp. 256–263).

Ein-Dor, P. (1999). Artificial intelligence: A short history and the next forty years. In K. E. Kendall (Ed.), *Emerging information technologies.* Thousand Oaks: Sage.

Erman, L., Hayes-Roth, F., Lesser, V., & Reddy, D. R. (1980). The hearsay II speech understanding system: Integrating knowledge to resolve uncertainty. *Computing Surveys, 12*(2), 213–253.

Gal, A. (1999). Semantic interoperability in information services: Experiencing with CoopWARE. *SIGMOD Record, 28*(1), 68–75.

Guha, R. V. (1991). *Contexts: A formalization and some applications*. Doctoral dissertation, Stanford University, Stanford, CT, USA (STAN-CS-91-1399-Thesis).

Hayes-Roth, B. (1985). A blackboard architecture for control. *Artificial Intelligence, 26*, 251–321.

Kahng, J., & McLeod, D. (1996). Dynamic classification ontologies for discovery in cooperative federated databases. *Proceedings of the First IFCIS International Conference on Cooperative Information Systems* (*CoopIS'96*), *Brussels, Belgium* (pp. 26–35). Belgium.

Lesser, V., Fennell, R., Erman, L., & Reddy, D. R. (1975). Organization of the Hearsay II speech understanding system. *IEEE Transactions on Human Factors in Electronics, ASSP-23*, 11–24.

McCarthy, J. (1987). Generality in artificial intelligence. *Communication of ACM, 30*, 1030–1035.

McCarthy, J., & Buvac, S. (1997). *Formalizing context, computing natural language* (pp. 13–50). Stanford: Stanford University.

Mena, E., Kashyap, V., Illarramendi, A., & Sheth, A. P. (2000). Imprecise answers in distributed environments: Estimation of information loss for multi-ontology based query processing. *International Journal of Cooperative Information Systems, 9*(4), 403–425.

Modica, G., Gal, A., & Jamil, H. M. (2001). The use of machine-generated ontologies in dynamic information seeking. *Proceedings of the Sixth International Conference on Cooperative Information Systems* (*CoopIS 2001*), *Trento*.

Motro, A., & Rakov, I. (1998). Estimating the quality of databases. *Lecture Notes in Computer Science, 1495*, 298.

Moulton, A., Madnick, S. E., & Siegel, M. (1998). Context mediation on wall street. *Proceedings of the 3rd IFCIS International Conference on Cooperative Information Systems* (*CoopIS'98*) (pp. 271–279). New York: IEEE-CS.

Ouksel, A. M., & Naiman, C. F. (1994). Coordinating context building in heterogeneous information systems. *Journal of Intelligent Information Systems, 3*(2), 151–183.

Papatheodorou, C., Vassiliou, A., & Simon, B. (2002). Discovery of ontologies for learning resources using word-based clustering. *Proceedings of the World Conference on Educational Multimedia, Hypermedia and Telecommunications* (*ED-MEDIA 2002*), *Denver, CO* (pp. 1523–1528).

Rijsbergen, C. J. (1979). *Information Retrieval* (2nd ed.). London: Butterworths.

Salton, G. (1989). *Automatic text processing: The transformation, analysis, and retrieval of information by a computer.* Reading: Addison-Wesley.

Salton, G., & McGill, M. J. (1983). *Introduction to modern information retrieval*. New York: McGraw-Hill.

Schuyler, P. L., Hole, W. T., & Tuttle, M. S. (1993). The UMLS (Unified Medical Language System) metathesaurus: Representing different views of biomedical concepts. *Bulletin of the Medical Library Association, 81*, 217–222.

Smith, H., & Poulter, K. (1999). Share the ontology in XML-based trading architectures. *Communications of the ACM, 42*(3), 110–111.

Soergel, D. (1985). *Organizing information: Principles of data base and retrieval systems*. Orlando: Academic.

Turney, P. (2002). *Mining the web for lexical knowledge to improve keyphrase extraction: Learning from labeled and unlabeled data*. (Tech. Rep. No. ERB-1096; NRC #44947). Washington, DC: National Research Council, Institute for Information Technology.

Valdes-Perez, R. E., & Pereira, F. (2000). Concise, intelligible, and approximate profiling of multiple classes. *International Journal of Human Computer Studies, 53*, 411–436.

Williams, T., Lowrance, J., Hanson, A., & Riseman, E. (1977). Model-building in the VISIONS system. *Proceedings of IJCAI-77, Cambridge, MA* (pp. 644–645).