GOOSE: A Goal-Oriented Search Engine
With Commonsense

Hugo Liu, Henry Lieberman, Ted Selker

MIT Media Laboratory
20 Ames St., E15-320G
Cambridge, MA 02139, USA
{hugo, lieber, selker}@media.mit.edu

Abstract. A novice search engine user may find searching the web for information difficult and frustrating because she may naturally express search goals rather than the topic keywords search engines need. In this paper, we present GOOSE (goal-oriented search engine), an adaptive search engine interface that uses natural language processing to parse a user's search goal, and uses "common sense" reasoning to translate this goal into an effective query. For a source of common sense knowledge, we use Open Mind, a knowledge base of approximately 400,000 simple facts such as "If a pet is sick, take it to the veterinarian" garnered from a Web-wide network of contributors. While we cannot be assured of the robustness of the common sense inference, in a substantial number of cases, GOOSE is more likely to satisfy the user's original search goals than simple keywords or conventional query expansion.

1 Introduction

The growth of available content on the World Wide Web makes it necessary for everyone to use tools, not experience, to find things. Major search engines like Google and Yahoo have made great progress in indexing a large percentage of the content on the web so that it is searchable. However, the user interface to the search process is usually just a text input box waiting for input. The user interfaces in most of today's search engines still rely on a grammar of set operators and keywords, and for good results, the user is expected to be able to fill the box with the right keywords and in the right combination. This situation prompts the question: instead of having the user conform to the search engine's specifications, why not make the search engine adapt to how the user most naturally expresses his/her information needs, so that even inexperienced users can perform an effective search?

1.1 An Experiment

To learn some qualities an intuitive search engine interface should have, we asked four search engine novices and four experienced search engine users to perform several tasks using the Yahoo search engine. Whereas experienced users chose precise

keywords likely to isolate the types of web pages they were looking for, novice users reverted to typing their search goal into the keyword field in natural language. For example, one search task that users were asked to perform was to find people on the web who shared the user's own interests. One novice user submitted the query: "I want to find other people who like movies," and obtained many irrelevant and unwanted search results on the topic of movies. In contrast, a more experienced user formed the query: " +'my homepage' +'my interests' +'movies' " and was able to get many relevant results. The experienced user chose not only a keyword ("movies") on the topic of the search, but also two keywords ("my homepage", "my interests") differentiating the context in which the topic keyword should appear. In choosing these keywords, the experienced user used her expertise to guide a series of inferences from the search goal. In interviewing the user, we learned that the inference chain, or thought process, that she went through looked something like this:

> *I want to find people online who like movies.*
> *Movies are a type of interest that a person might have.*
> *People might talk about their interests on their homepage*
> *People's homepages might start with "my homepage".*

This prompted us to further reasoning. As with all of the inference chains used by the four experienced users, this inference chain has the following property: Most of the steps in the inference chain are statements that arguably fall under the "common sense" knowledge domain, things that most people know to be true (in this case, only the last step is somewhat domain-specific knowledge); however, the knowledge of how to connect these commonsense facts to infer a good search query is where search engine expertise is required.

Even the experience of these few subjects point to out that novice searchers are confusing the search engine with the approach that they naturally communicate with. Simple improvements might be to: 1) allow the user to formulate the search query as a statement of the user's search goal, and from that, the search engine must make the necessary inference to arrive at the appropriate keywords, and 2) allow the user to express the search query in natural language.

To meet the second criterion, the search engine needs to have natural language parsing capabilities. The first criterion is trickier. If we are to expect the search engine to assume the burden of performing the inference, we might give it knowledge about the world that most people know (commonsense), and also some knowledge about what a good search query is, something that experienced search engine users know (expertise).

GOOSE is a goal-oriented search engine organized around the concept of a search goal. Enriched with commonsense knowledge, search engine expertise, and natural language parsing capabilities, it assumes the burden of translating a user's search goal into a good query.

In this paper, we will first present some background on this project, followed by descriptions of the GOOSE user interface and internal mechanism, a sample user scenario, and preliminary user test results. We will then proceed to discuss some future work of personalizing the commonsense and conclude.

2 Background

Previous approaches to query improvement have for the most part employed three techniques: 1) expanding the topic keyword using thesauri and co-occurrence lists [7, 10] 2) relevance feedback [11], and 3) using hand-crafted question templates [1]. Though the first approach shows promise for queries that return limited results, expanding keywords does not necessarily improve the relevance of the search results. The second approach does a better job of improving relevance, but complicates the task model by adding additional search steps. In addition, neither the first nor the second approaches address the weaknesses of keywords as the basis of the user interface. The third approach, as used by Ask Jeeves [1], offers the user a more intuitive natural language interface, but answerable questions must be anticipated in advance and a template for each question must be handcrafted. For this reason, we don't believe that this approach is easy to scale.

Our approach is significantly different from all the aforementioned approaches. In our system, the original query is a natural language statement of the user's search goal, and the reformulation step involves natural language parsing of this statement, followed by inference to generate the query that will best satisfy this goal. Unlike thesauri-driven keyword expansion, our system is not merely adding new keywords, but is actually performing inference and composing an entirely new search query that would best fulfill the user's goal. Compared with relevance feedback, the user interface we propose is automatic, and does not require additional steps in the task model. Finally, unlike handcrafted question templates, we believe that our approach of using a freely available, ever-growing, and vast source of commonsense knowledge to perform *reasoning* over the original query is a more scalable approach, and allows for many levels of inference, compared to the fixed, one-level of inference associated with question templates.

2.1 Source of Commonsense

The idea of using commonsense reasoning to improve user interfaces has been extensively explored by Minsky [5]. The commonsense knowledge used by our system comes from the Open Mind Commonsense Project [8] – an endeavor at the MIT Media Laboratory that aims to allow a web-community of teachers to collaboratively build a database of knowledge using diverse representations, and to explore ways to use this knowledge to make computer applications more intelligent and context-aware. Using the Open Mind Commonsense website, web collaborators input simple facts about the world, expressed as simple English sentences, which are organized into an ontology of commonsense relations.

2.2 Ordinary Commonsense vs. Application-Level Commonsense

When we refer to the commonsense knowledge used in GOOSE, we mean two things. The first is ordinary commonsense, which encompasses the things that people

normally consider to be known by everyone, such as "sugar tastes sweet," or "if someone hits you, you may feel pain." The second is application-level commonsense, that is, knowledge specific to a domain, and considered to be commonsense in that domain. An example of application-level commonsense in our web search engine domain is: "espn.com is a website which provides news about sports." Both types of commonsense can be easily solicited through the Open Mind website interface because each piece of knowledge is expressed in simple English. In addition, some application-level commonsense can be mined from the World Wide Web.
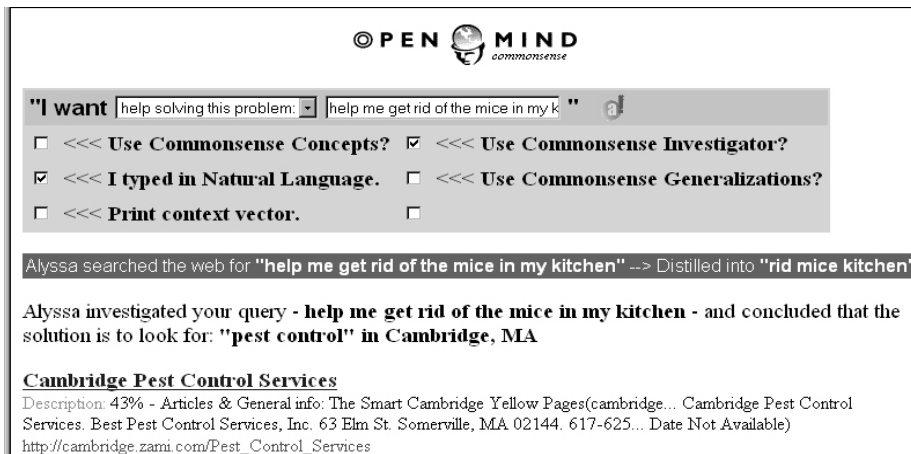
3   User Interface

Fig. 1. A screenshot of the current User Interface for GOOSE, where search goals must be manually disambiguated.

Arguably the most intuitive interface would simply allow the user to type the entire search goal as a sentence in natural language. Our system must then understand that goal and have the expertise to know how to reformulate the goal into a good query. Unfortunately, the expertise in our system is currently not complete enough to be able to interpret arbitrary goals, so instead we have created some templates that encapsulate search engine expertise for the common categories of goals.

GOOSE's user interface (Fig. 1) asks the user to select the goal of his/her search from a pull-down menu, and enter a query that completes the sentence begun by the search goal. Currently five search goals exist, and they are:

1. "I want help solving this problem:"
2. "I want to research…"
3. "I want to find websites about…"
4. "I want to find other people who…"
5. "I want specific information about the product/service…"

Because the knowledge associated with each goal category is modular, it would be relatively easy to add new search goals. Without an extensive usability study, it is unclear exactly how much coverage these categories of goals provide, and how many may eventually be needed; however, we believe that some of the categories listed above are generic enough to support any goal. Therefore, the issue of scaling up is not likely to be limited so much by the number and types of available search goals as by the diversity and coverage of the commonsense knowledge available for inference. One limitation of this type of interface is the lack of ability to state multiple goals and overlapping goals, but this is addressable if we allow more than one goal to be active at a time, and devise a method for combining search results obtained through multiple goals.

4 Mechanism

Given a search goal and search query, GOOSE performs four major internal steps before results are returned to the user: 1) Parsing the query into a semantic frame [6]; 2) classifying the query into a commonsense sub-domain; 3) reformulating the query through commonsense inference guided by expertise templates; 4) and re-ranking results using commonsense concepts.

4.1 Parsing to Frames

After the user executes the search query, GOOSE parses the query to fill the slots of a semantic frame, which provides a concise, stereotyped representation of the original query. Representing the original query as a frame makes commonsense reasoning easier because the most important features of the query are extracted. An example is given in Table 1.

Table 1. An example of a filled semantic frame for the goal, "I want help solving this problem:" and the query, "My cat is sick and wheezing"

Slot Name	Slot Value
Problem Attribute	[Sick, Wheezing]
Problem Object	[Cat]

As suggested by this example, each search goal needs its own unique set of semantic frames. This is true because different aspects of each query are useful to accomplishing different search goals. In the above example, identifying the problem attribute and problem object is most useful to identifying a solution through commonsense reasoning. The set of all semantic frame templates represents a part of the expertise that experienced users possess. Currently, each of the five search goal categories has one semantic frame, but as the system scales, more frames are likely to be added.

It is worth pointing out that our system's parsing of the natural language query differs from the ways in which other search engines handle unstructured text input. A typical search engine throws out a list of stop words and treats the remaining words as keywords [9], but our approach tokenizes, part-of-speech tags, parses the entire query, and translates the parse tree into a filled semantic frame.

4.2 Classification

In addition to parsing the original query to frames, a classifier examines the original search goal and determines the commonsense sub-domain that can provide the most applicable knowledge when performing commonsense reasoning. Examples of sub-domains for the "I want help solving this problem" search goal include "personal health problems," and "household problems." Each sub-domain contains both ordinary and application-level commonsense knowledge.

Classification is performed in a relatively straightforward way. Each sub-domain is described by the commonsense concepts it covers. A search goal is classified into all sub-domains, which match the concepts contained in it. Multiple sub-domain matches can be safely merged.

We have chosen to group the commonsense used for GOOSE into sub-domains to help disambiguate certain words and concepts. Another advantage of sub-domains of commonsense is the savings in the run-time of the inference, a benefit of a smaller search space.

4.3 Reformulation

In this step, we take the filled semantic frame and apply reasoning over the chosen commonsense sub-domain. In our current implementation, reasoning takes place as an inference chain, implemented as a depth-first search, guided by heuristically motivated rules that help direct the inference so as to avoid unnecessary searching. Inference terminates when an application-level rule has fired. Again, application-level rules are a component of search engine expertise. When the inference terminates, we will have the reformulated search terms that we need.

Once the query has been successfully reformulated, it is submitted to a commercial search engine and the result is captured for further refinement.

4.4 Re-ranking

Using GOOSE's concept vectors, a list of weighted words and phrases representative of the concepts contained within the search query, GOOSE re-ranks the search results so that the hits most relevant to the search query are given higher priority.

This concept-based re-ranking step is similar to the query expansion approach proposed by Klink [3], except that in our case, it is only used as a refinement of existing search results.

Where commonsense inference fails to infer any query from the search goal, keywords from the search goal are extracted and passed to the commercial search engine. In such cases, query refinement with commonsense concept vectors can serve as a back-up mechanism, because such refinement may still lead to improved results over the baseline where GOOSE is not used at all.

5   A Scenario

Having explained the GOOSE user interface and mechanism, let us imagine a typical user scenario. Suppose that a novice user has a sick pet and wants to find ways to remedy the problem. She does not know how to form a good search query, so she decides to try her search on GOOSE. She chooses the goal, "I want help solving this problem:" and types in the query, "my golden retriever has a cough." Using the semantic frame defined for this particular goal, GOOSE fills the frame as follows:

Problem Attribute: [cough]
Problem Object: [golden retriever]

The classifier examines the query and determines that the commonsense sub-domain to be used is "animals." Performing inference over the "animals" sub-domain, the following inference chain is invoked:

1. A golden retriever is a kind of dog.
2. A dog may be a kind of pet.
3. Something that coughs indicates it is sick.
4. Veterinarians can solve problems with pets that are sick.
5. Veterinarians are locally located.

The first three steps in the inference chain are ordinary commonsense, while the last two steps are application-level commonsense. GOOSE takes the result of the inference and submits the reformulated query, "Veterinarians, Cambridge, MA" to a commercial search engine. The locale that was added is a personalization, and must be obtained through a user profile. After search results are returned, commonsense concept vectors are used to refine the results so that search hits containing the concepts closest to "veterinarian" appear higher in the search results.

The user finds what she was looking for in the first page of results and never had to explicitly choose the keywords that brought her to what she was looking for.

6   Preliminary User Tests

We conducted preliminary user tests asking four search engine novice users to form queries for a few simple search tasks. Due to the limited search goal categories available in the current implementation, we focused on the categories and commonsense

sub-domains that the system knew how to handle. The query being inputted into the GOOSE UI was sent to both GOOSE as well as directly to the Google commercial search engine. Users were then asked to rate the relevance of the first page results on a scale of 1 (most irrelevant) to 10 (most relevant). In cases where commonsense inference failed to infer a search query from the search goal, commonsense concept vectors were still used to reorder the search results. Table 2 presents some of the results.

Table 2. Preliminary user test results. Participants formed 2 queries each for each search task.

Search Task	# successful inferences	Avg. score GOOSE	Avg. score Google
Solve household problem	7 / 8	6.1	3.5
Find someone online	4 / 8	4.0	3.6
Research a product	1 / 8	5.9	6.1
Learn more about	5 / 8	5.3	5.0

Our test results suggest that for novice search engine users, GOOSE on average produced more effective first-page results than Google, a leading commercial search engine. The problem solving goal category is where inference showed the most promising results, as demonstrated by the search task "solve household problem." However, the high rate of failure of the inference in producing a query suggests that GOOSE is still very brittle in the current implementation and only works well under very constrained domains for which organized commonsense knowledge exists, such as personal health and household problems. Many more domains and goals must be supported before any extensive user tests can performed, but these initial results are encouraging.

One fundamental limitation of using the commonsense knowledge for inference is illustrated by the result for the "research a product" search task. In this task, some users chose to search for trademark names of products, such as "the Total Gym", and "TurboTax". In such cases, commonsense inference will not be of help because trademark product names are not part of the knowledge base. This task, however, seems to be particularly suitable to a keywords approach, as Google received relatively high marks on this task. GOOSE received similar marks because although it could not be helpful in this case, it did not hurt the results either.

The results of this preliminary user test are promising. In future user tests, we hope to measure the intuitiveness of the proposed user interface, the usefulness of GOOSE to already experienced users, and have a head-to-head comparison against other query enrichment techniques such as keyword expansion and relevance feedback.

7 Conclusion

In this paper, we presented an adaptive search engine interface that can use common-sense to perform inference over user's search goal, in order to generate an effective query. While the commonsense inference is not complete, it can still be useful. GOOSE is a fail-soft application, in that, in the case it fails to produce a better query for the user, it will just produce the same results the user would have obtained anyway. So the argument can be made that GOOSE can be useful, even if the common-sense reasoning is brittle, because GOOSE can help some of the time, and it won't hurt the rest of the time.

As we continue to scale the commonsense coverage of GOOSE, we face several pointed issues. First, classification into commonsense sub-domains becomes less accurate as the number of sub-domains increases. Second, it will become increasingly difficult to define commonsense sub-domains that are of the right size and that do not overlap with existing sub-domains. One radical solution to the two above-mentioned problems would be to not only allow overlapping sub-domains, but to go so far as *fostering* many diverse and competing representations of commonsense, each with strengths and weaknesses, which will compete with each other in their reasoning of a particular query. In this model as suggested by Minsky [5], common-sense coverage will be increased, and reasoning will be more robust because it will exploit the complementary strengths of different representations.

Third, as the number of commonsense statements increase, inference will take combinatorially longer and be more prone to noise because the search space will have increased. To overcome these problems, we need to give the inference process more guidance via pruning techniques, and give it the ability to recognize when it is on the wrong path to the goal, or when it does not know enough to reach the goal. One possibility for guided inference is to valuate candidate inference chains that will result from different inference paths, much like a chess-playing program valuates board positions that result from different moves. However, this approach assumes that it is feasible to devise good ways to valuate an inference chain, which is a non-trivial problem.

From the preliminary user tests, we have learned more about the fundamental limitations of using commonsense to help the user compose search queries. First, the commonsense knowledge in Open Mind contains only about 400,000 facts, which do not assume equal distribution of the knowledge over diverse topics. Minsky estimates that somewhere on the order of 15 million pieces of knowledge may be needed in order to be comparable to what humans possess. Obtaining and organizing knowledge on that scale will be a huge challenge, not to mention efficiency issues that that scale creates. The second major limitation of commonsense is that it will probably not tell GOOSE about all the specific topics needed to perform inference, such as trademarked products, what specific companies do, etc. It can only help to reason about concepts and problems that we encounter in everyday life. Without speculation of the difficulty of doing so, if we can mine specific knowledge from other resources or the web, it may be possible to connect this knowledge to the inference mechanism.

8 Future Work

As of now, GOOSE is not yet robust or helpful enough although it has the potential to be. In addition to the scaling issues discussed above, we are working toward two goals: personalizing the commonsense, and automatic detection of goal categories.

One way to consider the role "common sense" plays in the system is to think of it as a generic user model, because it represents knowledge in everyone's head (everyone within a particular culture). We can customize the user model by adding personal commonsense to the system such as "Mary is my sister." Personalizing commonsense is logical, because the notion of what "common sense" is varies from one person to another. GOOSE may be able to utilize personal commonsense to better interpret the user's search goals and produce more relevant results. An example of personalization currently used by the system is the placement of the locale keyword in the query to accompany a local business. "Veterinarian Cambridge, MA" is one example. However, we can also imagine subtler examples of how personal commonsense can influence inference. For instance, if a user has a broken VCR, she might want search results for either do-it-yourself resources or electronics repair shops. Depending on the type of person she is, she might want one type of result or the other or both.

Personal commonsense can be stated as simple English sentences, so it is easy to add to the system. The real challenge is in devising a way to collect the information. For instance, we can imagine with the broken VCR example, that the user may be shown two sets of search results, and her preferring one set or the other set may then enter the appropriate piece of personal commonsense into the system. Other ways to enter personal commonsense can include an interview wizard, mining information about the user from a homepage, or getting shared information from some other context-aware application that is also learning personalizations about the user. ARIA [4], a photo agent also being developed at the MIT Media Lab, is an example of such application.

The second goal is to eliminate the explicit goal selection task by automatically classifying queries into goal categories. This may first require the coverage of the goal categories to be validated, and may necessitate more robust natural language processing in order to parse the unconstrained input. Alternatively, we may be able to apply shallow IR techniques such as support vector machines to perform the classification.

In the end, we hope to create a much more intuitive and personalized search experience for all web users, and to utilize the lessons learned here about commonsense reasoning so as to be able to apply its benefits to other domains and applications.

References

1. Ask Jeeves, Inc..: Ask Jeeves home page. (2002). http://askjeeves.com/.
2. Belkin, N.J.: Intelligent information retrieval: Whose intelligence? In: ISI '96: Proceedings of the Fifth International Symposium for Information Science. Konstanz: Universtaetsverlag Konstanz. (1996). 25-31.

3. Klink, S.: Query reformulation with collaborative concept-based expansion. Proceedings of the First International Workshop on Web Document Analysis, Seattle, WA (2001).

4. Lieberman, H., Liu, H.: Adaptive Linking between Text and Photos Using Common Sense Reasoning. In Proceedings of the 2nd International Conference on Adaptive Hypermedia and Adaptive Web Based Systems, Malaga, Spain (2002).

5. Minsky, M.: Commonsense-Based Interfaces. Communications of the ACM. Vol. 43, No. 8 (August, 2000), Pages 66-73

6. Minsky. M.: A Framework for Representing Knowledge. MIT, (1974). Also, In: P.H. Winston (Ed.): The Psychology of Computer Vision., McGraw-Hill, New York, (1975).

7. Peat, H. J. and Willett, P.: The limitations of term co-occurrence data for query expansion in document retrieval systems. Journal of the ASIS, 42(5), (1991), 378--383.

8. Singh, P.: The Public Acquisition of Commonsense Knowledge. AAAI Spring Symposium, Stanford University, Palo Alto, CA, (2002).

9. Shneiderman, B., Byrd, D., and Croft, B.: Sorting out searching: A user-interface framework for text searches, Communications of the ACM 41, 4 (April 1998), 95-98.

10. Voorhees, E.: Query expansion usin lexical-semantic relations. In Proceedings of ACM SIGIR Intl. Conf. on Research and Development in Information Retrieval. (1994) 61-69.

11. Xu, J. and Croft, W.B.: Query Expansion Using Local and Global Document Analysis. In Proceedings of the Nineteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, (1996). pp. 4-11.