

# Acquiring Knowledge about Human Goals from Search Query Logs

Markus Strohmaier\*

*Knowledge Management Institute, Graz University of Technology and Know-Center,  
Inffeldgasse 21a/II, A-8010 Graz, Austria*

Mark Kröll

*Knowledge Management Institute, Graz University of Technology and Know-Center,  
Inffeldgasse 21a/II, A-8010 Graz, Austria*

---

## Abstract

A better understanding of what motivates humans to perform certain actions is relevant for a range of research challenges including generating action sequences that implement goals (planning). A first step in this direction is the task of acquiring knowledge about human goals. In this work, we investigate whether Search Query Logs are a viable source for extracting expressions of human goals. For this purpose, we devise an algorithm that automatically identifies queries containing explicit goals such as `find home to rent in Florida`. Evaluation results of our algorithm achieve useful precision/recall values. We apply the classification algorithm to two large Search Query Logs, recorded by AOL and Microsoft Research in 2006, and obtain a set of  $\sim 110.000$  queries containing explicit goals. To study the nature of human goals in Search Query Logs, we conduct qualitative, quantitative and comparative analyses. Our findings suggest that Search Query Logs (i) represent a viable source for extracting human goals, (ii) contain a great variety of human goals and (iii) contain human goals that can be employed to complement existing commonsense knowledge bases. Finally, we illustrate the potential of goal knowledge for addressing following application scenario: to refine and extend commonsense knowledge with human goals from Search Query Logs. This work is relevant for (i) knowledge engineers interested in acquiring human goals from textual corpora and constructing knowledge bases of human goals (ii) researchers interested in studying characteristics of human goals in Search Query Logs.

*Keywords:* Knowledge Acquisition, Commonsense Knowledge, Human Goals,

---

\*Corresponding author. Tel.: +43 316 873 - 9257; Fax.: +43 316 873 - 9254  
*Email addresses:* `markus.strohmaier@tugraz.at` (Markus Strohmaier ),  
`mkroell@tugraz.at` (Mark Kröll)

## 1. Introduction

Knowledge about human goals has been found to be an important kind of knowledge for a range of challenging research problems. These challenges include goal recognition from people's actions, reasoning about people's goals or the generation of action sequences that implement goals (planning) (Schank & Abelson, 1977). Goal knowledge appears advantageous in practical problem settings as well, for instance, to inform companies to adjust their range of products or services according to their customers' desires.

Knowledge acquisition denotes a field of research concerned with gathering ontological constructs from experts or natural language corpora. Knowledge about human goals has been found to be important for a range of research communities and problems, including work on intelligent user interfaces (Smith & Lieberman, 2010), commonsense-enabled applications (Lieberman et al., 2004) and commonsense knowledge bases. Two ongoing research projects, Cyc (Lenat, 1995) and ConceptNet/Openmind (Liu & Singh, 2004; Singh et al., 2002), have been capturing commonsense knowledge, including knowledge about human goals, over the past years aiming to continuously refine, improve and extend their commonsense knowledge with different strategies. While Cyc partly relies on human experts to develop and build their knowledge base, ConceptNet aggregates and processes contributions made by volunteers all over the world.

These existing attempts illustrate two main problems in the process of constructing a knowledge base about human goals: 1) the goal acquisition problem (or bottleneck), which refers to the costs associated with knowledge acquisition (Lieberman et al., 2007) and 2) the goal coverage problem, which refers to the difficulty of capturing the tremendous variety and range in the set of human goals (Eslick, 2006). These problems have hindered progress in capturing broad knowledge about human goals. To address the goal acquisition problem, Search Query Logs referred by (Battelle, 2005) to as *Databases of Intentions* - appear to be an interesting candidate for this task.

In this work, we are interested in the extent to which we can tap into Search Query Logs, to extract expressions of human goals in an automated way. Intuitively, Search Query Logs appear to represent an appropriate corpus for this task because every query can be considered to be an expression of a person's (search) intent. In recent years, considerable progress was made in approximating a person's search intent for the purpose of improving web search. The focus of our work is different: While previous research on search intent focused in part on categorizing queries into high-level goal taxonomies thereby serving a functional purpose (e.g. to improve search, cf. (Baeza-Yates et al., 2006; Jansen et al., 2008; Lee et al., 2005)), we know little about the acquisition of individual human goal expressions from Search Query Logs for knowledge acquisition purposes (as in (Jansen, Booth, Park, Zhang, Kathuria & Bonner, 2009) or (Liu

Table 1: Exemplary queries from two introduced query classes are contrasted: Queries containing explicit goals and queries which do not contain explicit goals. The presented exemplary queries were obtained from the AOL Search Query Log.

<i>Queries containing explicit goals</i>	<i>Queries not containing explicit goals</i>
“sell my car”	“Mazda dealership”
“play online poker”	“online games”
“find home to rent in Florida”	“Miami beach houses”
“passing a drug test”	“drug test”
“raising your credit score”	“credit cards”

et al., 2002)). While our research is related to query log analysis (cf. for example (Broder, 2002; Jansen et al., 2008; Lee et al., 2005; Murray & Teevan, 2007; Pasca et al., 2007)), our goal is to contribute to the problem of commonsense knowledge acquisition.

### 1.1. Queries containing Explicit Goals

In most cases, a query submitted to a search engine expresses some user’s underlying goal or motivation. While some goals contained in search queries might be very explicit, other queries might contain more implicit goals, which would mean that they are more difficult to recognize by, for example, an external observer. To give an example: in terms of intentional explicitness, the query `car Miami` differs from the query `buy a car in Miami` (Strohmaier et al., 2007). This observation suggests that it is useful to distinguish between at least two classes of queries: (1) queries that contain explicit goals and (2) queries which do not. Table 1 contrasts queries of both classes.

Results from a larger human subject study corroborate the existence of these two classes and furthermore hint towards a theoretical separability (Strohmaier et al., 2008). To tap into Search Query Logs for knowledge acquisition purposes, we propose to automatically identify and extract queries which contain explicit goals. Our algorithm learns characteristics of queries containing explicit goals in order to perform the classification task. By applying the algorithm to two large Search Query Logs recorded by AOL and Microsoft Research in 2006, we obtain a set of  $\sim 110.000$  unique queries that contain explicit goals.

To study the acquisition of knowledge about human goals extracted from Search Query Logs, we address following four research questions<sup>1</sup>:

- **RQ 01:** Do queries containing explicit goals exist in Search Query Logs?
- **RQ 02:** How accurately can we identify queries containing explicit goals?
- **RQ 03:** What are characteristics of queries containing explicit goals automatically extracted from Search Query Logs?

---

<sup>1</sup>We elaborate on these research questions in Section 3.2.

- RQ 04: Do Search Query Logs contain commonsense goals, i.e. goals that are found in ConceptNet, a commonsense knowledge base? If they do, what is the nature of human goals shared by ConceptNet and Search Query Logs and how do they differ?

Our findings suggest (i) that Search Query Logs represent a viable source for extracting expressions of human goals, (ii) that Search Query Logs contain a great variety of human goals and (iii) that this variety has the potential to complement commonsense goals found in existing commonsense knowledge bases. To illustrate the benefits of acquiring goal knowledge, we employ this knowledge to following practical scenario: We outline how commonsense knowledge can be refined and extended by using human goals from Search Query Logs.

The overall contribution of this paper is the introduction of Search Query Logs as a viable, yet largely untapped, source for the task of knowledge acquisition with regard to human goals. This work is particularly relevant (i) for knowledge engineers interested in constructing knowledge bases of human goals and (ii) for researchers interested in studying characteristics of human goals in Search Query Logs. In applications, goal knowledge can benefit various domains including text analysis, human computer interaction or planning. The potential of goal knowledge to inform human computer interaction is already being investigated (cf. (Lieberman, 2008)). By equipping user interfaces with knowledge about human goals (Faaborg & Lieberman, 2006; Smith & Lieberman, 2010), a better understanding about users' actions can be achieved to advance the vision of more intelligent user interfaces.

## 2. Related Work

To the best of our knowledge, this work is novel in a sense that it studies the potential of Search Query Logs as source for goal knowledge acquisition. In the following, we first provide a brief survey of Information Extraction(IE) from textual resources since IE can be considered key technology to knowledge acquisition. Second, we review work from two related but previously unconnected fields of research, (i) Commonsense Knowledge Acquisition (including human goals) and (ii) Search Intent Detection and Categorization.

### 2.1. Information Extraction from Text

Information Extraction (IE) refers to the automatic extraction of structured information from unstructured textual resources. It also refers to the process of making information explicit and thereby useable by both humans and machines; information that otherwise remains hidden in vast amounts of digital text documents, e.g. newspaper archives. Extracted information is often stored in relational databases where the information becomes accessible, e.g. to querying mechanisms. To better understand the functions IE systems should perform, consider following sentence: *Vienna is the stylish capital of Austria.* An IE system takes this sentence as input and is expected to output a mapping to a

relational tuple, for instance, (“Vienna”, “CapitalOf”, “Austria”). The process of accurately extracting tuples from text requires already existing knowledge, i.e. a certain degree of human involvement. This condition has been termed *knowledge engineering bottleneck*.

The advent of the World Wide Web triggered a paradigm shift in IE due to changing requirements: lesser human involvement given the sheer amount of digital resources. Traditional IE approaches had to evolve towards automatic procedures. In the following, we will briefly survey this evolution; the chronological grouping was partly adopted from (Etzioni et al., 2008).

### 2.1.1. Traditional Information Extraction

Traditional Information Extraction (IE) systems focused on locating instances of narrow, pre-specified relations, such as the time and place of events, from small, homogeneous corpora. The spectrum of relation types was continually enlarged to support, for instance, constructing lexical databases such as WordNet (Hearst, 1998). (Hindle, 1990) and (Hearst, 1992) were among the first who pioneered hand-crafted, textual patterns in the early 1990’s. For their analyses, they took into account the surrounding context, e.g., syntactical and grammatical characteristics. While Hindle focused on predicate-argument structure, Hearst developed lexico-syntactic patterns, i.e. part-of-speech enriched regular expressions, to extract hyponymy (“is-a”) relations from text. Being hand-crafted, only instances of predefined relation types, e.g. hyponymy, were extracted. Additional disadvantages of traditional IE included domain dependency as well as being heavily time-consuming which led to research in automatic IE techniques.

### 2.1.2. Automatic Information Extraction

The objective of automatic Information Extraction (IE) is to continuously reduce human involvement in the IE process. As it turned out, machine learning was among the preferred methods and ideally complemented pattern based approaches. Human involvement was still key in order to provide learning algorithms with annotated training examples. Yet, instead of crafting patterns by hand, researchers attempted to automatically learn these patterns to reduce human efforts. First work in this direction included Soderland’s CRYSTAL system (Soderland et al., 1995), Kim’s PALKKA system (Kim & Moldovan, 1995), and Riloff’s AutoSlog-TS system (Riloff, 1996). In contrast to CRYSTAL and PALKKA, the AutoSlog-TS system represented an unsupervised approach to IE, i.e., the system generated extraction patterns using untagged text.

The annotation process proved to be a major bottleneck to highly scalable IE systems. To minimize human involvement, (Agichtein & Gravano, 2000) presented the IE system *Snowball* which extracted structured data from plain-text documents. The idea is to provide few but frequent training examples in combination with a regular expression the examples have to match. Snowball uses iteration cycles to repeatedly check the quality of the extracted instances. These cyclic quality checks reduce error propagation and therefore represent the main advancement compared to Brin’s Dual Iterative Pattern Expansion (DIPRE)

algorithm (Brin, 1999).

Self-supervised IE systems can be regarded as a subcategory of unsupervised methods. Yet, unlike classic unsupervised methods, self-supervised IE systems find and annotate examples on their own to train a classifier. Representatives include *KnowItAll* (Etzioni et al., 2005), a domain-independent system, that automatically extracts information from the web. KnowItAll is seeded with an extensible ontology and a small number of generic rule templates from which it creates text extraction rules for each class and each relation in its ontology. In (Etzioni et al., 2005), the authors also provided detailed information on lessons-learned towards the development of an Open Information Extraction (Open IE) system.

The term *Open IE* was coined by (Banko et al., 2007) and represents a novel extraction paradigm; the paradigm is meant to address challenges of extracting information from web-scale corpora: Open IE does not require (i) domain specific training data, (ii) in advance specification of relations to extract, (iii) but does require linear scalability due to massive data amounts. *TextRunner* (Banko et al., 2007) represents a fully implemented Open IE system which features all previously introduced requirements. When compared to its predecessor Know-ItAll (Etzioni et al., 2005), TextRunner’s average error rate is significantly lower while identifying an almost identical number of correct extractions. In addition, TextRunner extracts information from all relations at once thereby drastically reducing processing time.

The concept of “never-ending” (Carlson et al., 2010) or “life-long” (Banko & Etzioni, 2007) learning developed simultaneously. It refers to a 24/7 effort to continuously extract information from the web. In comparison to Open IE, the emphasis lies on constructing a comprehensive reflection of the web’s factual content; predicates to learn are given in advance. (Carlson et al., 2010) presented *NELL*, an implementation of a never-ending language learning system. The system consists of four subsystem components which utilize semi-supervised learning methods thereby simultaneously attempting to extract candidate facts. A component called knowledge integrator is then responsible for upgrading high-confidence candidates to the status of beliefs. First evaluations of NELL yielded promising precision results while constantly accumulating knowledge.

As is the case with recent IE methods, our work uses supervised machine learning to automatically extract expressions of human goals from Search Query Logs focusing on “hyponymy” relations. Feature engineering provides us with discriminant features for the classification task. However, these features reflect merely query log characteristics and thus cannot simply be transferred to other textual corpora such as weblogs or tweets.

## 2.2. Commonsense Knowledge Acquisition

Knowledge acquisition from natural language text often focuses on gathering a specific kind of knowledge such as “commonsense knowledge”. Commonsense knowledge spans a broad spectrum of human experiences such as *a lemon is sour*

or *if you forget someone’s birthday, they may be unhappy with you*. Since it is assumed that every person possesses commonsense, it is in general omitted from social conversations. Commonsense comprises fact-based knowledge as well as knowledge about other aspects including emotional aspects, temporal contexts or human goals.

To humans, “commonsense knowledge” appears trivial since it states simple facts about the world and its people. Yet, to make knowledge about the world accessible to machines, it needs to be acquired and adequately represented. A common approach to storing and structuring knowledge is to construct a knowledge base. Knowledge in structured form represents a prerequisite so that reasoning and inference mechanism can be applied. Reasoning, for instance, helps to answer *why* questions and can thus support, e.g. intelligent agents in their decision making processes.

CyC (Lenat, 1995) or ConceptNet/Openmind (Liu & Singh, 2004; Singh et al., 2002) are research projects aiming to capture commonsense knowledge, including knowledge about human goals. ConceptNet was designed to make practical context-based inferences over real-world texts. ConceptNet’s internal structure can be described as set of triples where each triple consists of two concepts connected by a semantic relation. ConceptNet provides a set of over 20 semantic relations which cover various thematics such as affectionate (MotivationOf), causal (EffectOf) or events (SubeventOf). Table 2 shows a list of ConceptNet entries taken from (Liu & Singh, 2004).

Table 2: In this table, a list of ConceptNet entries is shown. Each entry is represented as a triple consisting of a right and a left concept which are connected by a semantic relation.

Left Concept	Semantic Relation	Right Concept
eat breakfast	SubeventOf	read newspaper
eat breakfast	EffectOf	full stomach
wake up in the morning	SubeventOf	eat breakfast
wake up in the morning	Generalisation	wake up

The knowledge acquisition process itself may adopt a number of different strategies, including human knowledge engineering (Lenat, 1995), volunteer-based (Singh et al., 2002), game-based (Lieberman et al., 2007; Law et al., 2009) or semi-automatic approaches (Eslick, 2006). In (Singh et al., 2002), the authors described the Open Mind Experiences (OMEX) system which aimed to gather descriptions and explanations of everyday, “commonsense” experiences in form of stories. Telling stories represents a human trait to share knowledge, a circumstance the OMEX system took advantage of.

Knowledge about human goals has been found to be an important kind of knowledge for a range of challenging research problems, such as goal recognition from user actions, reasoning about human goals, or the generation of action sequences that implement goals (planning)(Schank & Abelson, 1977). We refer to the task of acquiring goals from textual resources as *Goal Mining*. This

problem covers a broad range of interesting aspects, including the acquisition of goals from scientific articles (Hui & Yu, 2005), organizational policies (Potts et al., 1994), organizational guidelines and procedures (Liaskos et al., 2006), Search Query Logs (Strohmaier et al., 2008) and others.

In the area of understanding natural language text, knowledge about human goals gains significance as a novel dimension to be considered. (Passonneau & Litman, 1993) theoretically analyzed whether the task of text segmentation can be informed by employing knowledge about peoples' goals. (Tatu, 2005) analyzed human goals in natural language text to improve the task of question answering. Extracting expressions of human goals to complement social media monitoring tools has been recently explored by (Kröll & Strohmaier, 2009). In this previous work, we studied political speeches from a goal-oriented perspective and classified human goals into a human goal taxonomy (Chulef et al., 2001). Eventually, we were able to compare political speeches not only by traditional topic category distributions but also by human goal category distributions.

Knowledge about human goals has been found to play a fundamental role in explanation, justification, and rationalization as well. Understanding peoples' goals can help to answer *why* questions about user behavior and user interactions (Faaborg & Lieberman, 2006; Lieberman et al., 2007; Smith & Lieberman, 2010). In commonsense enabled applications (Lieberman, 2008), explicit representations of goal knowledge are crucial for plan recognition and planning. In addition, they are an enabler for intelligent user interfaces which exhibit traits of commonsense understanding, such as goal-oriented search (Liu et al., 2002) or goal-oriented event planning (Smith, 2007).

### 2.3. Search Intent Detection and Categorization

The main objective of research in this field is concerned with estimating a searcher's intent to inform and improve the search process. Several definitions of search intent, e.g. (Broder, 2002), (Rose & Levinson, 2004) or (Downey et al., 2008), emerged each incorporating different aspects during search.

(Broder, 2002) introduced a high level taxonomy of search intent by categorizing search queries into three categories: navigational, informational and transactional. Two years later, (Rose & Levinson, 2004) started another attempt to search intent categorization. They repeatedly revised their goal categories based on empirical evidence. Their efforts resulted in a search intent hierarchy where high-level categories resembled Broder's taxonomy. Follow up research led to evolutions of Broder's work which included collapsing categories, adding categories (Baeza-Yates et al., 2006) and/or focusing on subsets only (Lee et al., 2005). In the context of goal knowledge acquisition, we can make two observations: First, definitions of search intent do not require a query's intent to be made explicit. Second, the adopted abstraction level suggests that information about individual human goals, e.g. **how to sell my car**, is lost. In contrast to Broder's understanding of search intent, we do not incorporate high-level categories of search intent but rather focus on individual human goals (e.g. informational vs. **buying a car**). (Downey et al., 2008) interpreted the information seeking process differently: They proposed to use subsequent actions



that succeeded a query as characterizations of the searcher’s goal. The last URL visited in a search session served as proxy for the user’s search intent. While their approach is useful to study user behavior with regard to high-level search intent, it cannot easily be used for extracting plausible human goals.

Complementary research was conducted by Jansen et al. who investigated search and search intent as decision making process (Jansen & McNeese, 2005) and consecutively as learning process (Jansen, Booth & Smith, 2009). In the latter, they analyzed the search process from a cognitive learning perspective. In their studies, they utilized a cognitive learning taxonomy (Anderson & Krathwohl, 2001) consisting of six categories: Remembering, Understanding, Applying, Analyzing, Evaluating and Creating. This taxonomy represents another potential high-level categorization scheme of search intent, e.g. Does the user intend (i) to recall something from memory (Remembering) or (ii) to collect information for a decision (Evaluating).

A few publications in recent years were dedicated to algorithms that automatically categorized queries according to their high-level search intent. (Kang & Kim, 2003), (Lee et al., 2005) and (He et al., 2007) exploited statistical and syntactical features of Search Query Logs, e.g. user-click behavior or part-of-speech information, to perform the categorization task. (Beitzel et al., 2005) proposed a method for automatic query classification by leveraging unlabeled data within a semi-supervised learning framework. Their semi-supervised approach facilitated the augmentation of labeled training samples for the classification task. (Baeza-Yates et al., 2006) applied a combination of supervised and unsupervised learning to classify queries. Besides categorizing queries according to their intent, the attempt was made to capture the user’s interest as well. This was accomplished by assigning queries to top-level categories of the Open Directory Project. Learning more about the relation between intent versus topic was the focus in (Jansen & Booth, 2010). The authors showed that intent categories varied across topics and presented a real-time algorithm that took this observation into account. In (Jansen et al., 2008), the authors operationalized Broder’s categories and provided a detailed description of (i) the process and (ii) the metrics used to classify queries accordingly. (Li et al., 2008) described a semi-supervised learning approach to query intent classification with the use of search click graphs. Based on the distances in this click graph, the authors infer intent classes of unlabeled queries from those of labeled ones. In (Hu et al., 2009), the authors’ motivation for putting effort into understanding user intent was to identify an adequate vertical search engine. By mining Wikipedia’s link structure, they were capable of generating classification models for a wide range of intent classes such as “travel” or “job” intent. In a departure from statistical approaches, (Guo & Agichtein, 2010) investigated whether human computer interactions during search such as mouse movements supported search intent inference more accurately. Their work was motivated by the observation that similar queries exhibited different underlying search intents.

Areas such as e-commerce and personalization also realized potential benefits of approximating a person’s search intent. In e-commerce, (Dai et al., 2006) proposed detecting commercial intent to determine whether a user expresses

commercial interest (such as purchase, selling or auction) or not. In personalization, (Teevan et al., 2008) examined the variability of goals across different users. (Li et al., 2008) focused on specific aspects of user intent such as job intent or product intent. (Yi et al., 2009) explored ways to predict geographic intent, if it is not stated explicitly such as the query “Disney world”. By taking into account geographic information during the search procedure, they showed that retrieval performance could be improved.

In this paper, we aim to contribute to research on knowledge acquisition, in particular, acquiring knowledge about human goals. While we do not aim to contribute to search intent detection and categorization, our work might inspire new ways of employing knowledge of human goals in intelligent systems and intelligent user interfaces (cf. (Strohmaier et al., 2009)). We believe that *tapping into Search Query Logs for knowledge acquisition purposes* represents a unique problem in the context of goal knowledge acquisition, and – to the best of our knowledge – has not been studied before.

### 3. Research Design

In this section, we provide a detailed description (i) of the datasets used in our experiments and (ii) of the research methodology we employ to address the research questions.

#### 3.1. Datasets

##### *Search Query Logs from AOL & MS Research:*

In this work, we employ two large Search Query Logs which were recorded by AOL and Microsoft Research in 2006. We combine these two Search Query Logs to (i) increase the number of queries as well as (ii) to decrease potential domain and population bias that is introduced by using only one Search Query Log. To give an example, queries such as `cancel AOL account` or `how to delete the msn account` reflect a certain degree of domain bias. The first query log, the MSN Search Query Log<sup>2</sup> excerpt, contains ~15 million queries (from US users) that were sampled over one month in May, 2006. The second log, the AOL Search Query Log (Pass et al., 2006), contains ~20 million queries (from US users) recorded between March 1, 2006 and May 31, 2006. Search queries from both logs were extracted using the same method, and underwent several sanitization and pre-processing steps in order to reduce noise to an acceptable level:

---

<sup>2</sup>The MSN Search Asset Data Spring 2006 represents a data set which was provided by Microsoft for selected paper proposals at the Workshop on Web Search Click Data 2009 in conjunction with WSDM’09 (<http://research.microsoft.com/en-us/um/people/nickcr/wscd09/>) accessed Feb 28th, 2011.

- *Empty Queries*: We removed blank queries and queries containing just a minus character.
- *Short Queries*: We restricted ourselves to queries with at least three tokens ( $n > 2$ ) for the following two reasons: (i) inherent ambiguity of short queries and (ii) the lack of syntactical structure to express human goals. This restriction resulted in a removal of  $\sim 65\%$  of the queries contained in the original datasets.
- *URL queries*: We removed queries containing URLs or fragments of URLs using regular expressions.
- *Queries containing lyrics or movie titles*: In preliminary experiments, we observed that queries for music lyrics (I **need love lyrics**) often contained a verb, but referred to songs rather than actual human goals. This bears the risk of confusing our classification approach that is in part based on syntactic features. However, such queries can be identified, since they often contain keywords such as “lyrics” or result in click-through to lyrics or movie related websites (e.g. “http://www.seeklyrics.com”). We performed limited term and website blacklisting to heuristically reduce the number of such queries in the datasets.
- *Syntax check*: We removed queries containing tokens, which are not numbers or sequences of letters. We used this filter to eliminate corrupted character encodings.
- *Removed misspellings*: We removed misspelled queries. Whether or not a consecutive query represents a spelling correction was determined by the Levenshtein distance (Levenshtein, 1966) between two consecutive query strings. A query was removed if the Levenshtein distance between the query and its successor is  $\leq 2$  and the first query has no click-through event attached.

By applying these filtering steps, we only use  $< 5\%$  of queries in the query log. While this reduction appears rigorous, we point out that only this part of all search queries is of value to this work (from a knowledge acquisition as opposed to an information retrieval perspective).

In this work we had only access to two time-constrained Search Query Logs, i.e. from AOL and from Microsoft Research. As a consequence, we cannot assume that our findings generalize for Search Query Logs in general, across time and space. Query logs from other search engines, e.g. Google, might exhibit different properties which would need to be studied in future work.

#### *Commonsense Knowledge obtained from ConceptNet:*

In this work, we choose ConceptNet (Liu & Singh, 2004) as commonsense knowledge base because of its open availability, its natural language knowledge representation and its considerable size. Moreover, knowledge in ConceptNet is

partly represented in free-form text which facilitates the comparison with search queries. In this work, we regard knowledge which is contained within ConceptNet, as commonsense knowledge. We identify commonsense goals (Lieberman et al., 2007) in ConceptNet by querying concepts (ConceptNet nodes) which are connected by relations such as *MotivatedByGoal*, *UsedFor* and *CapableOf*. We compiled a subset of entries from ConceptNet that consists of commonsense goals and imposed the following restrictions on all entries: Commonsense goals had to contain at least one verb and at least one noun. To enforce this restriction, we examined corresponding part-of-speech tags<sup>3</sup>. For our experiments, we obtained a set of  $\sim 68.000$  commonsense goals from ConceptNet.

### 3.2. Research Methodology

In the following, we briefly summarize how we address each research question:

#### *RQ 01: Do queries containing explicit goals exist in Search Query Logs?*

To address the first research question, we conduct a human subject study where human annotators manually classify 3000 search queries into two classes: queries containing an explicit goal and queries which do not. Similarly to previous work (cf. (Broder, 2002) or (Rose & Levinson, 2004)), the manual classification task allows us to develop a better understanding of the class' characteristic and their boundaries. This study corroborates our intuition that Search Query Logs contain a small percentage of human goal expressions; the distribution of the human subjects' annotations hints towards a theoretical separability of these two classes.

#### *RQ 02: How accurately can we identify queries containing explicit goals?*

We start by inspecting characteristics of both query classes that can serve as potential features for our classification approach. Inspired by related work from search intent classification (cf. (Kang & Kim, 2003; Lee et al., 2005; Dai et al., 2006)), we initially take a wide spectrum of feature types into account including click-through as well as part-of-speech information. We apply feature selection methods to decide on a final set of feature types that appear promising for the classification task. Having identified these features, we apply two established classification models, i.e., Naive Bayes (Friedman, 1997) and Support Vector Machines (Dumais et al., 1998; Vapnik, 1998). To evaluate these models, we perform three-fold cross validation and calculate standard metrics: precision, recall and F1-measure.

#### *RQ 03: What are characteristics of queries containing explicit goals automatically extracted from Search Query Logs?*

---

<sup>3</sup>Stanford Part-Of-Speech Tagger version 1.6 available from <http://nlp.stanford.edu/software/tagger.shtml> accessed Feb 28th, 2011.

We conduct quantitative as well as qualitative analyses to learn more about the nature of human goals acquired from Search Query Logs. We refer to work from query log analysis such as (Pass et al., 2006; Jansen et al., 2000; Jones & Diaz, 2007) which analyze different aspects of queries, e.g., temporal aspects. While evaluation strategies appear to be similar, the intention behind query log analysis often is to improve retrieval performance. Our findings provide us with insights into the nature of frequent and infrequent human goals. To learn more about their diversity, we analyze verbs in human goals by classifying them into selected Levin’s verb classes (Levin, 1993).

*RQ 04: Do Search Query Logs contain commonsense goals, i.e. goals that are found in ConceptNet, a commonsense knowledge base? If they do, what is the nature of human goals shared by ConceptNet and Search Query Logs and how do they differ?*

By addressing these research questions, we aim to establish a connection between human goal expressions from two domains, i.e. commonsense and search; we regard connecting these two domains and corresponding insights, e.g. common and uncommon characteristics, as acquiring knowledge about human goals.

We approach these questions by first verifying the existence of commonsense goals in Search Query Logs. Second, we study characteristics of commonsense goals by generating verb class histograms of selected Levin’s verb classes (Levin, 1993). By learning more about common and uncommon features, we develop a better understanding of how human goals from Search Query Logs could contribute to complementing commonsense knowledge. As motivation to enrich existing taxonomic structures with automatically extracted in stances, we selectively refer to (Pantel & Pennacchiotti, 2008) and (Suchanek et al., 2007). Pantel et al. attempt to link extracted instances into WordNet (Fellbaum, 1998). Suchanek et al. present YAGO, a light-weight and extensible ontology. For ontology creation, they use Wikipedia to automatically extract facts and attempt to unify them afterwards with WordNet. Similarly, we believe that human goals from Search Query Logs have the potential to contribute to commonsense knowledge. In Section 5, we present an illustration scenario as a first step into this direction.

## 4. Results

### *4.1. RQ 01: Do queries containing explicit goals exist in Search Query Logs?*

In this subsection, we introduce a practical definition to identify queries containing explicit goals. While a multitude of definitions for human goals exist in related literature, our definition seeks to be applicable in the context of Search Query Logs, i.e., capable of distinguishing search queries that contain explicit goals from queries which do not. Based on work that emphasizes the crucial role of verbs in explicit goals (Liaskos et al., 2006; Regev & Wegmann, 2005), we define queries containing explicit goals in the following way:

*A search query is regarded to contain an explicit goal whenever the query 1) contains at least one verb and 2) describes a plausible state of affairs that the user may want to achieve or avoid (cf. (Regev & Wegmann, 2005)) 3) in a recognizable way (cf. (Strohmaier et al., 2008)).*

“Recognizable” refers to what (Kirsh, 1990) defines as “trivial to identify” by a subject within a given attention span. According to Kirsh, “trivial to identify” represents the ability to make a decision in constant time. To give an example: The query **passing a drug test** recognizably reflects the searcher’s goal. In contrast, the query **drug test** allows several interpretations of the searcher’s underlying intent. “Plausible” refers to an external observer’s assessment whether the human goal contained in a query could likely represent the goal of a user who formulates the given query. With regard to the previous example, we can generally assume that **passing a drug test** represents a plausible goal. A counterexample is represented by the query **living on the moon**. It is important to note that it would be rather difficult to completely verify this assessment solely based on data from an anonymous query log due to the inherent goal verification problem of such a task (Strohmaier et al., 2008). However, the objectives of our work are more modest: we are interested in acquiring *plausible* human goals for *knowledge acquisition* purposes. An advantage of acquiring broad knowledge about *plausible* human goals is that it can put constraints on the space of *possible* human goals, which plays a role in, for example, goal recognition (He et al., 2007) or query disambiguation (Allan & Raghavan, 2002).

A query does not contain an explicit goal when it is difficult or extremely hard to elicit some specific goal from the query. Examples include blank queries, or queries such as **car** or **travel**, which embody search goals on a very general, ambiguous and mostly implicit level.

To explore the agreeability of our definition and the feasibility of an automatic approach, we conducted a human subject study in which 4 judges (Computer Science graduate students) were instructed to annotate a small query sample. In this task, the judges conducted a question answering task; they were required to independently answer a single question for each of 3000 queries randomly obtained from the AOL Search Query Log<sup>4</sup>. The question for each query followed this schema:

*Given a query X, do you think that Y (with Y being the first verb in X, plus the remainder of X) is a plausible goal of a searcher who is performing the query X? Two examples should illustrate the process:*

---

<sup>4</sup>The MSN query log was not available at the time the agreeability study was conducted.

---

Given query: “how to increase virtual memory”

Question: Do you think that “increase virtual memory” is a plausible goal of a searcher who is performing the query “how to increase virtual memory”?

Potential Answer: Yes

Given query: “boys kissing girls”

Question: Do you think that “kissing girls” is a plausible goal of a searcher who is performing the query “boys kissing girls”?

Potential Answer: No

---

After the question answering task, we assigned the answers for each query to the corresponding categories in the following way: each answer “Yes” resulted in classifying the query as a “query containing an explicit goal”; each answer “No” resulted in classifying the query as a “query not containing an explicit goal”. The chart in Figure 1 shows that 243 queries out of 3000 have been labeled as containing an explicit goal by all 4 subjects (8,1%, right most bar), and 134 queries have been labelled as containing an explicit goal by 3 out of 4 subjects. This shows that 1) Search Query Logs contain human goals and 2) the number of queries containing human goals is expected to be small<sup>5</sup>. We would like to emphasize the dichotomous characteristic of the agreement distribution in Figure 1, which provides preliminary evidence for (i) the agreeability of our construct and (ii) the potential for an automatic classification approach.

To further explore agreeability, we calculated the inter-rater agreement  $\kappa$  (Cohen, 1960) between all pairs of human subjects A, B, C and D. The  $\kappa$  values in our human subject study range from 0,65 to 0,76 (see Figure 1). The average inter-rater agreement  $\kappa$  yields  $\sim 0,72$  which hints towards a principal (yet not optimal) agreeability of our definition.

#### 4.2. RQ 02: How accurately can we identify queries containing explicit goals?

To devise an algorithm to automatically identify queries containing explicit goals, we compiled a training data set that was based on a majority vote among the participants of the human subject study presented in the previous subsection. Out of the 3000 labeled queries, the negative examples were defined by the two bars on the left hand side of Figure 1 (2525 total), and the positive examples were defined by the two bars on the right hand side (377). The bar in the middle represents controversial queries<sup>6</sup> which were removed. Altogether, our training set for the classification task comprised 2902 queries.

We considered several feature types for our automatic classification approach including “Plain Text”, “Part-of-Speech Trigrams”, “Query Length”, “Click-Through” and “Language Modeling”. After preliminary evaluation cycles, we

---

<sup>5</sup>Note that we use a query subset in our human subject study. If we took filtered queries into account as well, the fraction of queries containing human goals would be even smaller.

<sup>6</sup>Controversial queries are queries where the majority of judges do not agree whether the query contains an explicit goal or not. Controversial queries could include ambiguous as well as unambiguous queries which do not contain explicit human goals.

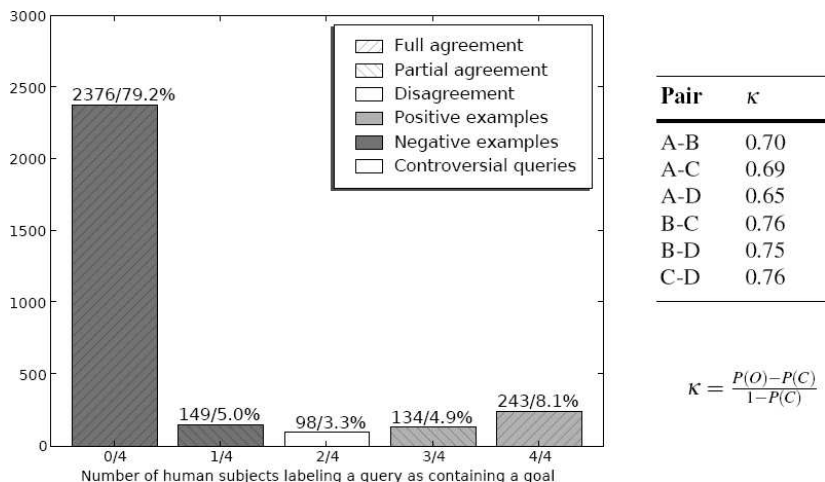


Figure 1: (left) shows the percentage of human subjects who labeled a given query as containing an explicit goal. We calculated the inter-rater agreement  $\kappa$  between all pairs of human subjects A, B, C and D to further explore agreeability (see table on the right). An average inter-rater agreement  $\kappa$  of 0,72 indicates good (but not optimal) agreeability of our definition. The formula for calculating  $\kappa$  is denoted in the right lower corner where  $P(O)$  denotes the relative observed agreement among raters and  $P(C)$  denotes the hypothetical probability of chance agreement, i.e. if the two raters were totally independent.

decided on using the first two feature types since they exhibited sufficient discriminative power for our classification task at hand:

- Plain Text: Queries are represented as binary word vectors or “Set of Words” (SoW). The Porter stemming algorithm (Porter, 1997) was used for word conflation and removing stop words.
- Part-of-Speech Trigrams: Each query is translated from a sequence of tokens into a sequence of part-of-speech (POS) tags. For the part-of-speech tagging, we used a Maximum Entropy Tagger<sup>7</sup> that had been trained on sections 0 to 18 of the Wall Street Journal part of the Penn Treebank corpus. Trigrams were generated by moving a fixed sized window of length 3 over the POS sequence. The sequence boundaries were expanded by introducing a single marker (\$) at the beginning and at the end allowing for length two POS features. The query “buying/VBG a/DT car/NN” would yield the following trigrams:

\$ VBG DT; VBG DT NN; DT NN \$

Our intuition behind introducing trigrams was to exploit the grammatical structure of explicit goal queries, i.e. putting emphasis on verb phrases.

<sup>7</sup>[http://homepages.inf.ed.ac.uk/lzhang10/maxent\\_toolkit.html](http://homepages.inf.ed.ac.uk/lzhang10/maxent_toolkit.html) accessed Feb 28th, 2011.



Throughout our experiments, we used WEKA (Witten & Frank, 2005) as data mining toolkit for feature pre-processing, feature selection, classification and evaluation of classification models. We experimented with several feature types such as word n-grams, part-of-speech n-grams or query length. By ranking these features according to the results of a chi-square feature selection we determined most discriminative features which eventually led to the decision to use only word unigrams and part-of-speech trigrams. Table 3 lists the 20 most discriminative features together with example queries for each feature and the number of occurrences of the feature in the positive class (#).

We would have expected that word features such as “how” and “where” were amongst highranking features to identify queries containing human goals. We suspect a (probably too strict) stop word removal by the Porter stemmer to be responsible for this absence. However, the value of indicators like “how” and “where” to the classification task is preserved in some highly ranked part-of-speech trigrams, i.e. “\$ WRB TO” or “WRB TO VB”. Moreover, we can observe that only a fifth of the features in Table 3 are unigram features, notably all of them verbs. Thus, it appears that the most discriminative features for identifying queries containing explicit goals are POS features complemented by verbs.

Table 3: The top 20 most discriminative features are illustrated resulting from applying WEKA’s chi-square feature selection. The part of the search query that matches the respective part-of-speech feature is enclosed by brackets ([ ]). To aid readability, descriptions of selected part-of-speech tags are provided according to the Penn Treebank Tag Set: *WRB* represents a Wh- adverb, *VB* represents the base form of a verb, *PRP* represents a personal pronoun, *NN* represents a noun in singular form, *IN* represents a preposition and *DT* represents a determiner.

Nr.	#	Feature	Example Matching Query	Nr.	#	Feature	Example Matching Query
1	126	\$ WRB TO	[\$ where to] find shrooms in Georgia	11	12	TO VB PRP	how [ to copyright your] photos
2	130	WRB TO VB	[how to live ] jewishly	12	14	WRB VBP PRP	my hair turned orange [how do I ] fix it
3	83	TO VB NN	drink milk [to lose weight]	13	26	TO VB NNS	what [to pay Mexicans]
4	41	buy	buy acoustic guitar	14	28	VB NN NNS	[make business cards]
5	58	VB NN NN	[find property values ] calculator	15	19	TO VB DT	teach yourself [to play the] piano
6	20	find	find an old friend for free	16	9	VB PRP JJ	how to [get yourself sick]
7	36	TO VB JJ	I want [to download instant] messenger	17	45	TO VB IN	places [to stay in] Gatlinburg
8	27	make	make your own parable	18	8	install	install Microsoft windows 2000
9	52	\$ VB NN	[\$ find lawyer ] in Georgia to form llc	19	14	\$ VB PRP	[\$ customize your] aol buddy icon
10	29	VB NN IN	[ borrow money from ] Donald Trump	20	22	VB PRP NN	how to [obtain us passport]

After having identified a set of discriminative features, we apply two common classification models for handling textual data, i.e. a Naive Bayes (NB) classifier (Friedman, 1997) and a linear Support Vector Machine (SVM) (Dumais et al.,

1998; Vapnik, 1998). Similar to prior work on query classification (Li et al., 2005), we choose the F1 measure, i.e. the harmonic mean of precision and recall, for evaluation. Since we are mainly interested in achieving high values for the positive class, i.e. queries containing explicit goals, we only report precision, recall and F1 values for the positive class.

In conducting experiments with regard to the F1 score, we aim to identify configurations that balance precision and recall in a way that is useful for acquiring expressions of human goals. We evaluate the selected linear classification models with regard to varying feature set sizes (see Figure 2). Feature sets are generated by applying WEKA’s chi-square feature selection and keeping the top  $N$  features. For each classifier/feature set size combination, 10 trials of three-fold cross-validation are carried out. The resulting scores for each combination are averaged over all trials. Figure 2 presents two resulting learning curves, i.e. the F1 scores of different feature set sizes and classification models.

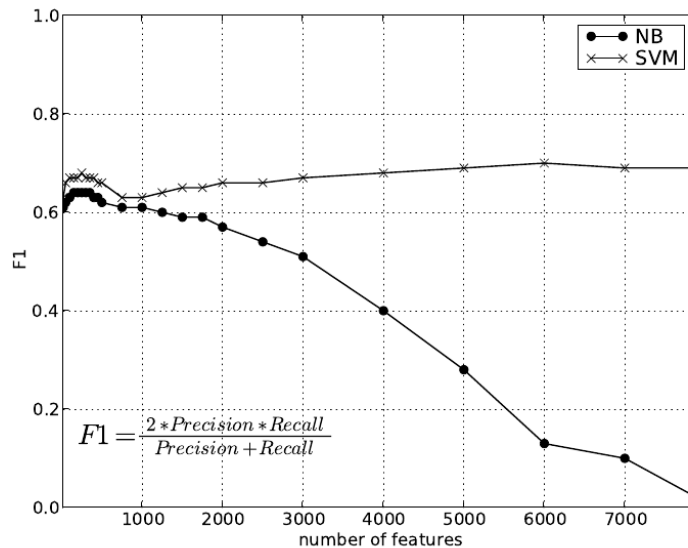


Figure 2: Learning curves for two classification models, i.e. the linear Support Vector Machine and the Naive Bayes classifier are shown. F1 measures indicate the respective classifier’s performance with varying feature numbers. F1-scores refer to the positive class, i.e. queries containing explicit goals.

These results indicate that the SVM appears to be better suited for our classification task, in particular with increasing number of features. The results also illustrate the NB classifier’s dependence on accurate feature selection prior to training and classification. For the NB classifier, the performance significantly deteriorates when more features are used. Informed by these results, we select the linear SVM as the classification model for our subsequent experiments.

Table 4: Following averaged Precision, Recall and F1-scores are achieved on the manually labeled data set. A linear SVM takes into account all features to conduct the classification task. Precision, recall and F1 scores refer to the positive class, i.e. queries containing explicit goals.

Precision	Recall	F1 – measure
0.77	0.63	0.69

Table 5: Classification results are shown in form of a confusion matrix, i.e. reflecting true positive (TP), false positive (FP), false negative (FN) and true negative (TN) information. The table provides an overview of the query distribution as well as corresponding query examples. Examining exemplary queries categorized as FP or FN can be beneficial to better understand the algorithm’s behavior and to improve its performance.

Classified as → Annotated as ↓	Containing an Explicit Goal	Not Containing an Explicit Goal
Containing an Explicit Goal	TP: # 239 / 8,2% Query Examples: “how to write a resume”, “make money from home”, “obtaining a passport”	FN: # 138 / 4,8% Query Examples: “flyfishing around park city Utah”, “lowering home insurance ”
Not Containing an Explicit Goal	FP: # 73 / 2,5% Query Examples: “dancing with the stars”, “stem cell research”, “living room furniture”	TN: # 2452 / 84,5 % Query Examples: “online games”, “national car rental”, “vegas ride shuttle”, “shoppers guide”

Table 4 shows the precision, recall and F1 scores for the positive class, i.e. queries containing explicit goals. The values result from averaging 10 trials of three-fold cross-validation keeping all features. A precision of 77% means that in 77% where the classification model believed the query contained a goal, the majority of human subjects agreed. This form of evaluation allows statements about the generalization capabilities of our algorithm: A precision of 77% is the quality to expect when applying our algorithm to Search Query Logs. While this precision score is comparable to other attempts to acquire commonsense knowledge (cf. ConceptNet), it reflects our approach’s limitations and requires further attention by future research.

Table 5 shows classification results from our automated method in form of a confusion matrix. It provides an overview of the query distribution regarding true positive (TP), false positive (FP), false negative (FN) and true negative (TN) information. In addition to frequency values, Table 5 provides corresponding query examples. Examining exemplary queries categorized as FP or FN can be beneficial to better understand the algorithm’s behavior as well as to improve its performance. As will be discussed in Section 4.3, incorrectly classified entries are mainly due to incorrect part-of-speech tagging.

As simple baseline approach, we would guess that a query containing a verb always contains an explicit goal. Yet, such a baseline would perform significantly worse: While the baseline would excel on recall (= yielding a recall of 1,0 due

to our definition of explicit goals requiring a query to contain a verb), it would perform worse with regard to the extraction task due to low precision. In our experiments, the baseline achieved a precision of 0,13 and a F1 score of 0,23.

#### *4.3. RQ 03: What are characteristics of queries containing explicit goals automatically extracted from Search Query Logs?*

To address this question, we conduct quantitative as well as qualitative analyses to gain more insights into the nature of human goals which we automatically extracted from Search Query Logs. We apply our automatic classification method (see Section 4.2) to  $\sim 35$  million queries<sup>8</sup>, i.e. the AOL and the MSN Search Query Log combined. The set of queries our system classified as containing human goals, which we call the result set, comprises  $\sim 142.000$  queries, 110.000 of which are unique. With a precision of 77%, this means an estimated 109.000 queries in the result set actually do contain goals. 109.000 queries might appear small in the light of  $\sim 35$  million queries contained in the original Search Query Logs. Yet, in case of the AOL Search Query Log, 20 million queries reportedly represent only 0,33% of the total number of queries served during that time. Considering the large numbers of queries served every day, the approach would be able to continuously extract human goal expressions on larger datasets.

Stop word removal and stemming were applied to the result set to obtain a more accurate frequency ranking. Similar entries such as **buy a new car** and **buying new cars** are then merged into one entry **buy car** with higher frequency values. To enhance readability, all queries in our result figures and tables are manually post-processed: stems are manually extended to their base form and, if necessary, stop words are re-inserted to restore original meaning.

### **Quantitative Analysis**

The 40 most frequent queries from the result set are presented in Table 6. Each example is accompanied by rank and frequency information. Queries containing the token **http** are filtered out and those queries containing expletives or sex-related content are replaced by **deleted**.

The information in this table reflects – to some extent – the needs and goals of the North-American web population, i.e. users of AOL/MS search during the period of the dataset recordings. Some of the most frequent queries containing human goals relate to commonsense goals such as **lose weight**, **get pregnant** or **listen to music** (marked with a \* in Table 6). We refer to these goals as commonsense goal because of their relation to ConceptNet, a commonsense knowledge base. **lose weight** represents a ConceptNet node which is connected to other nodes by relations such as **MotivatedByGoal**. The existence of commonsense goals among the most frequent queries in the result set provides some evidence that Search Query Logs are suited for the task of commonsense

---

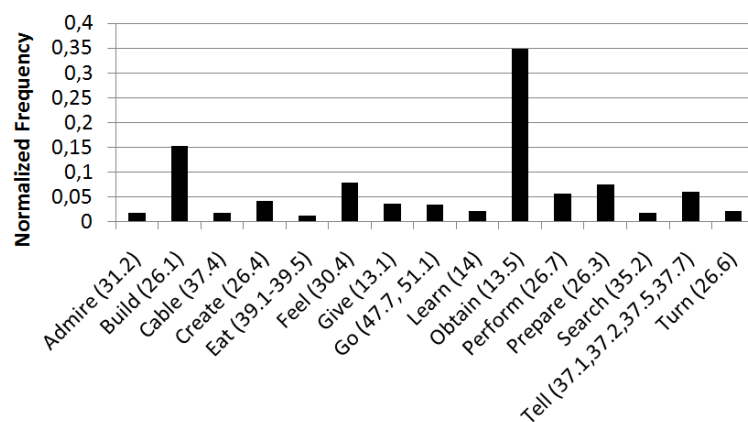
<sup>8</sup>After initial filtering,  $\sim 1,7$  million queries remained. This means that we only use  $<5\%$  of queries in the query log.

Table 6: The 40 most frequent queries containing goals in the result set are listed. Queries marked with a \* represent queries that are contained in ConceptNet indicating the existence of commonsense goals in Search Query Logs. Stems are manually extended to their base form and, if necessary, stop words are re-inserted to restore original meaning.

Nr.	Query	Freq.	Nr.	Query	Freq.	Nr.	Query	Freq.	Nr.	Query	Freq.
1	enterprise rent car	311	11	dollar rent car	142	21	buy buy baby	98	31	deleted	78
2	build bear	212	12	rent car	142	22	find people free	97	32	hertz rent car	73
3	pimp ride	195	13	find person*	139	23	find grave	96	33	lose weight*	73
4	rent center	192	14	find email address	138	24	listen free music	84	34	trick truck	70
5	listen to music*	190	15	tie ties	111	25	make money home	83	35	buy house*	67
6	find phone number*	185	16	meaning of name	109	26	start own business	83	36	work home*	66
7	assist sell	173	17	change password	107	27	write resume	83	37	lose weight fast*	64
8	pimp space	167	18	find address*	103	28	flash rack	81	38	make own website	63
9	budget rent car	166	19	pimp myspace	102	29	cancel aol account	78	39	play guitar*	63
10	find zip code	154	20	deleted	102	30	get pregnant*	78	40	gain weight*	62

knowledge acquisition. By combining the AOL and MSN Search Query Logs, we were able to partly decrease bias that would be introduced by using just one dataset. Yet, the remaining bias introduced by the corpus itself (search queries) and the population (i.e. AOL and MSN users) deserves attention: A fraction of frequent queries deals with web-related or AOL/MSN specific issues, such as the queries `find e-mail address` or `cancel aol account`. Entries such as `meaning of name`, and `buy buy baby` likely represent false positives, revealing two kinds of shortcomings of our approach: First, the automatic classification approach relies on linguistic patterns generated by part-of-speech (POS) tagging. In case of the query `meaning of name`, the POS tagger mistakenly tagged `meaning` as a verb (VBG) which yields an incorrect decision. Other examples include `enterprise rent car` and `hertz rent car` where the word `rent` has again been mistakenly tagged as a verb. A part-of-speech tagger that is trained on a more suitable corpus might help alleviating such problems in the future. Second, certain queries containing explicit goals resemble book titles, TV shows or music themes such as `buy buy baby`. This problem could be addressed by including domain knowledge (for instance “imdb.com”) in the classification task or inspecting and analyzing click-through data and anchor text, which can be expected to improve the overall performance of our approach.

To better understand the nature of identified human goals, we conduct a term analysis by identifying the 10 most frequent nouns and verbs in the result set and analyzing verb/noun co-occurrences. The most popular verb/noun co-occurrences in Table 7 seem to be indicative of typical human goals on the web, such as `make money`, `listen music` or `find phone`. Preliminary evaluations of



Selected Levin's Verb Classes

Figure 3: This histogram illustrates the distribution of explicit goal queries from the long tail over selected Levin's verb classes. The distribution is diverse with regard to all 15 verb classes (except predominance of verb classes "Build" and "Obtain") and thus provides evidence that the majority of queries containing explicit goals are diverse in nature. Frequency values are normalized to allow better comparison. Levin's verb classes along with corresponding indices are denoted in brackets.

Table 7: The 10 most frequent verbs and the 10 most frequent nouns in the result set are illustrated as well as frequency numbers of corresponding co-occurrences. Frequency values marked with a \* represent queries that are contained in ConceptNet.

	home (2541)	car (2227)	account (1729)	card (1537)	house (1365)	music (1322)	money (1276)	phone (1109)	name (1109)	window (1086)
<b>make</b> (12600)	294*	55*	24	318*	23*	97*	769*	31*	85	59*
<b>buy</b> (11905)	426*	457*	15	139	339*	84	74	154	17	57
<b>find</b> (11836)	231*	127*	26	33	133*	65*	72*	560*	260	41
<b>get</b> (7260)	102*	51*	35*	75*	84*	36	66*	38*	42	33
<b>do</b> (4767)	52	22*	25*	40	21	19	33	21	22	42
<b>listen</b> (2864)	14	0	0	0	2	702*	0	6	1	7
<b>use</b> (2490)	26	145*	24	22*	9	7	15*	52*	14	33
<b>clean</b> (2485)	29*	17*	2	1	72*	0	3	0	0	28*
<b>build</b> (2410)	108*	46*	15	3	128*	2	5	0	0	7
<b>write</b> (1899)	3	3	0	38*	0	14*	10	1*	16*	3

the top verb/noun correlations reveal that many of these human goals are also contained in the ConceptNet commonsense knowledge base (marked with a \*). This can be understood as a further indicator of the usefulness of Search Query Logs for acquiring knowledge about human goals. It also suggests that Search Query Logs might be useful to automatically complement knowledge contained in existing commonsense knowledge bases, which has been attempted before (Eslick, 2006).

If Search Query Logs would be utilized for such a purpose, a relevant question to ask is: *How diverse is the set of human goals contained in Search Query Logs?* The diversity of goals would ultimately constrain the utility of a given dataset for complementing existing knowledge bases. In order to explore this question, we classified explicit goal queries from the long tail (Anderson, 2004) into selected Levin’s verb classes (Levin, 1993). The histogram in Figure 3 reflects the classification result. This distribution provides evidence that the majority of queries containing explicit goals are diverse in nature, i.e. covering a broad spectrum of goals.

### Qualitative Analysis

While the analyses conducted so far provide statistical insights into the nature of human goals contained in search queries, it is difficult to infer information about their quality. To address this issue, we perform limited qualitative analyses through inspection. We select four verbs and four nouns, manually chosen to represent a range of exemplary activities and topics typically addressed by web search users and inspect explicit goal queries which contain them. In Table 8, the 10 most frequent goals in the result set are listed, which contain either the verbs **get**, **make**, **change** or **be**. Frequency refers to the occurrence in the result set. The human goals listed in Table 8 are the result of identifying the first verb in a query containing a goal, and truncating any tokens prior to this verb. Goals marked with a \* represent goals that are contained in ConceptNet. Many entries in Table 8 are related to existing commonsense goals, such as **be pregnant**, **be rich** or **be funny**.

To gain further insights, we attempted to select interesting nouns which belong to Information Extraction classes: **money** [economy], **birth** [event], **home** [location] and **people** [person]. The corresponding top 10 most frequent human goal expressions are depicted in Table 9.

Table 8 and Table 9 make an interesting case for using Search Query Logs to complement existing commonsense knowledge bases as we will demonstrate in Section 5.

*4.4. RQ 04: Do Search Query Logs contain commonsense goals, i.e. goals that are found in ConceptNet, a commonsense knowledge base? If they do, what is the nature of human goals shared by ConceptNet and Search Query Logs and how do they differ?*

By addressing this research questions, we aim to establish a connection between human goal expressions from two domains, i.e. commonsense and search.

Table 8: In this table, we provide an overview of the 10 most frequent human goals from Search Query Logs containing the verbs **get**, **make**, **change**, or **be**. Human goals marked with a \* indicate goals that are also contained in ConceptNet. Stems are manually extended to their base form and, if necessary, stop words are re-inserted to restore original meaning.

Nr.	Verb: get	Verb: make	Verb: change	Verb: be
1	get pregnant* (78)	make money home (83)	change password (107)	be anorexic* (27)
2	get rid of ants (45)	make your own website (63)	change screen name (47)	be funny* (11)
3	get passport* (43)	make money online (56)	change name* (31)	be bulimic (11)
4	get rid of love handles (23)	make wish foundation (53)	change aol password (28)	be cool* (8)
5	get out of debt* (22)	make money* (52)	change profile (20)	be loved (8)
6	get rid of stretch marks (20)	make the band 3 (49)	change aol screen name (17)	be millionaire* (8)
7	get myspace school(20)	make money fast (42)	change e-mail address (15)	be sexy* (6)
8	get rid of moles (19)	make new screen name (33)	change home page (14)	be emo (6)
9	get rid of belly fat (16)	make crossword puzzle (31)	change welcome screen (11)	be romantic* (6)
10	get rich* (15)	make ethanol (30)	change life* (7)	be happy* (5)

Table 9: In this table the 10 most frequent human goals from Search Query Logs containing the nouns **money**, **birth**, **home**, or **people** are presented. Human goals marked with a \* indicate goals that are also contained in ConceptNet. Stems are manually extended to their base form and, if necessary, stop words are re-inserted to restore original meaning.

Nr.	Noun: money	Noun: birth	Noun: home	Noun: people
1	make money from/at home (83)	find birth mother (6)	make money from/at home (83)	find people for free (97)
2	make money online (56)	get birth certificate (6)	work at home*(66)	find people (16)
3	make money* (52)	find birth parents (6)	buy home* (54)	find peoples phone number (12)
4	make money fast (42)	buy birth control pills (5)	sell home (35)	find missing people (9)
5	save money* (23)	get birth control (3)	build own home (31)	find peoples address (9)
6	make money on ebay (18)	buy birth control online (3)	design own home (24)	find people online (8)
7	ways to make money (17)	buy birth control (3)	sell own home (21)	win friends and influence people (8)
8	make money on internet (16)	use birth control pill (3)	find home* (18)	deal with difficult people (8)
9	find lost money (10)	obtain birth certificate (2)	make msn home page (16)	loop up people (6)
10	invest money* (9)	look up birth parents (2)	organize home* (16)	search for people (6)

We approached these questions by comparing human goals acquired from Search Query Logs to commonsense goals (cf. (Lieberman et al., 2007)) from ConceptNet. We identified commonsense goals in ConceptNet by querying concepts (ConceptNet nodes) which were connected by relations such as *UsedFor*, *CapableOf* and *MotivatedByGoal*. We compiled a subset of entries from ConceptNet that consists of commonsense goals and imposed the following restrictions on all candidates to increase quality: Commonsense goal candidates had to contain at least one verb and at least one noun.

Using this approach, we obtained an overall number of ~68.000 commonsense goals and compared them to the ~110.000 unique human goals acquired



from Search Query Logs. First, we were interested whether Search Query Logs contained commonsense goals. As an approximation, we calculated the intersection between Search Query Log and ConceptNet goal sets with the following intuition: An adequate number of shared entries would indicate the presence of commonsense goals in Search Query Logs. To calculate the intersection, we devised a simple goal matching algorithm (similar to (Lieberman et al., 2007)) to identify matching pairs of human goals from the two goal sets. All entries were pre-processed: stop words were removed and all remaining tokens were stemmed using the Porter stemmer (Porter, 1997). Table 4.4 shows examples of commonsense goal examples from ConceptNet before and after applying pre-processing steps.

Table 10: This table illustrates commonsense goal examples from ConceptNet before and after pre-processing.

Before Pre-Processing	After Pre-Processing
to protect your family	["protect", "famili"]
have something to do during breakfast	["have", "do", "breakfast"]
how to tell kids about suicide	["tell", "kid", "suicid"]

For two goals to match, they had to contain an equal number of identical stems. Our algorithm focused on lexical characteristics only; semantic similarities were not taken into account. This idea is similar to Liu’s process of normalization to identify similar instances in ConceptNet (Havasi, 2007). Table 11 illustrates some examples of matching and non-matching entries from ConceptNet and Search Query Logs.

Table 11: This table presents examples for matching and non-matching ConceptNet and Search Query Log entries. Matching sequences are highlighted.

ConceptNet Goals	Search Query Log Goals	Match
<u>make paper airplanes</u>	how to <u>make paper airplanes</u>	yes
<u>get in shape</u>	<u>getting into shape</u>	yes
we <u>buy houses</u>	<u>buy a house</u>	yes
we buy houses	purchase a house	no
<u>make money</u>	<u>make more money</u>	yes
make money	make money online	no

Eventually, we obtained ~2300 ConceptNet goals and ~3100 Search Query Log goals (occurrences) that produced positive matches. While the number of shared goals appears small, our findings provide first evidence of the existence of commonsense goals in Search Query Logs.

To answer the second question, we categorized the set of ~3100 commonsense Search Query Log goals into a subset of Levin’s verb class taxonomy (Levin, 1993). From this taxonomy, we selected 15 verb classes (see Figure 4) that we deemed relevant for reflecting human activities, e.g. **Eat** or **Learn**. Figure

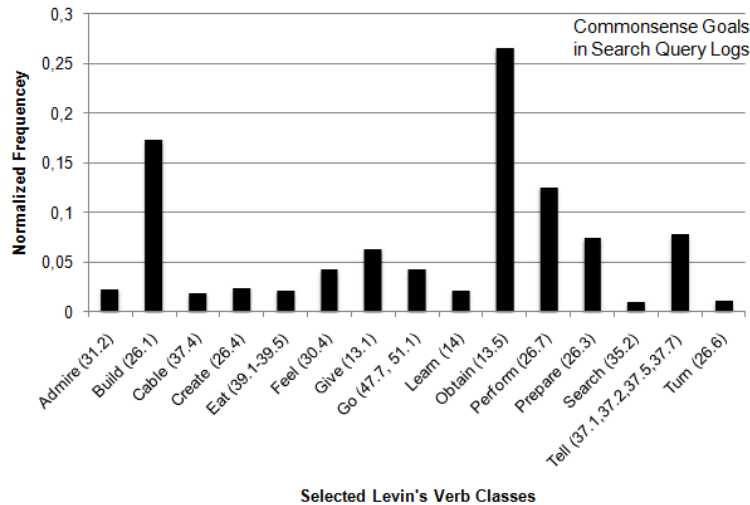


Figure 4: This histogram illustrates how commonsense goals in Search Query Logs are distributed over selected Levin’s verb classes. Frequency values are normalized to allow better comparison. Levin’s verb classes along with corresponding indices are denoted in brackets.

4 shows the resulting verb class histogram of commonsense goals from Search Query Logs.

Three dominant verb classes **build**, **obtain** and **perform** can be identified from Figure 4. Their dominance might be explained by occurrences of verbs in frequently stated commonsense goals such as **make money**, **buy food** or **play an instrument**. Classes **build** (verb **make**), **obtain** (verb **buy**) and **perform** (verb **play**) represent corresponding Levin’s verb classes. An interesting observation is that the verb class **search** is not dominant in our dataset. We believe this class is underrepresented due to the fact that search engines already represent a means for searching the web, i.e. goals in search queries do not need to contain verbs expressing the goal to search itself. In the following, we will study to what extent commonsense goals which are contained in Search Query Logs differ to those contained in ConceptNet.

To illustrate different characteristics between goals from Search Query Logs(QL) and ConceptNet(CN), we generate a verb class histogram for the complementary sets of goals. In set-theoretic terms, this reads as follows:  $CN - QL$  and  $QL - CN$ . Our initial intuition was that ConceptNet’s commonsense goals would be biased towards everyday situations and human characteristics such as *eating*, *feeling* and *living*. The results confirm our intuition:

The verb histogram in Figure 5 shows that verb classes **eat**, **gorge**, **touch**, and **feel** are more prominent in the ConceptNet set. Similarly, we expected classes such as **obtain** to be the dominating Search Query Log goals. This can be observed in our results: Levin’s verb class **obtain** dominates the human

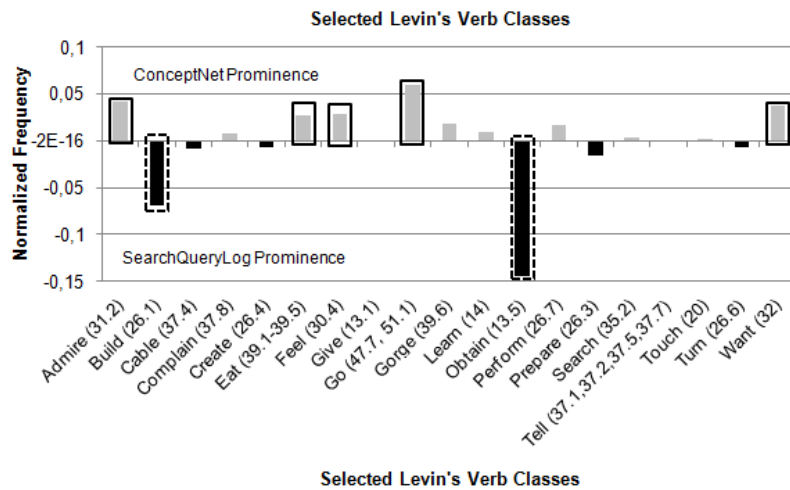
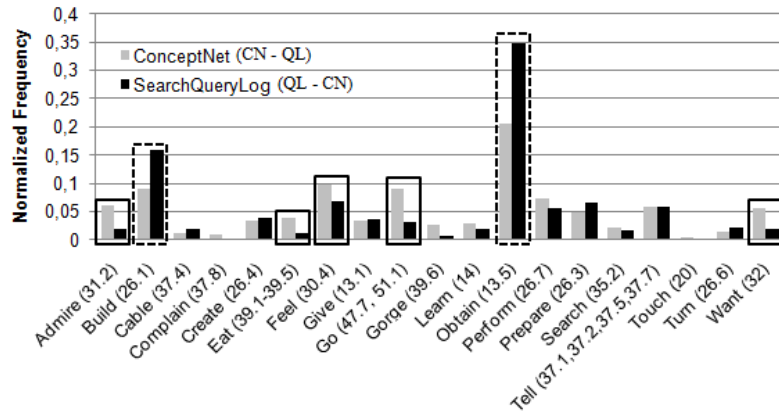


Figure 5: The top figure compares verb class distributions from two complementary sets, i.e. CN – QL and QL – CN. This comparison illustrates for which classes Search Query Logs can potentially contribute to increase coverage of ConceptNet. The bottom figure represents a different presentation of the upper figure's frequency values. It shows relative differences in frequency values to emphasize verb class prominences. Frequency values are normalized to allow better comparisons. Levin's verb classes along with corresponding indices are denoted in brackets.

goals acquired from Search Query Logs, which contain frequently occurring verb instances such as **get**, **buy** and **find**.

In addition, Figure 5 reveals that each dataset, i.e. CN – QL and QL – CN, favors a different range of human activities. In fact, this suggests that query logs could actually *help increase coverage of commonsense knowledge bases*, for example by focusing on types of goals that are more prevalent in Search Query Logs, such as **obtain** and **build**. The distribution also suggests that Search Query Logs are not suited to contribute to commonsense goals from verb classes such as **eat** or **feel**.

We conclude this subsection by making following three statements: 1.) Search Query Logs appear to be a potential source for commonsense goals. 2.) Search Query Logs and ConceptNet each emphasize different goal classes. 3.) Search Query Logs might represent a useful resource to complement existing commonsense knowledge bases such as ConceptNet which we will elaborate on in the next section.

## 5. Illustration Scenario of Applying Goal Knowledge: Complementing A Commonsense Knowledge Base

In this section, we aim to make use of our findings from Section 4.4 and sketch potentials of applying knowledge about human goals to complement ConceptNet, a commonsense knowledge base. We explore two methods for complementing ConceptNet: (i) refinement and (ii) extension. The internal structure of ConceptNet facilitates refinements and extensions simply by adding novel triples consisting of two concepts connected by a semantic relation.

First, we propose refining existing nodes by making them more specific, e.g. by adding adjectives or adverbs. The ConceptNet goal **finding friends** can thus be refined by following selected Search Query Log goals **find old friends**, **find a lost friend** or **find free military friends**. Table 12 provides potential refinements for a set of commonsense goals from ConceptNet.

In addition, Search Query Logs provide us with frequency information that is denoted in brackets for each potential refinement in Table 12. A quality criterion for refinement candidates could, for instance, be to introduce a frequency threshold.

Second, we demonstrate one possible way to extend commonsense knowledge by human goals acquired from Search Query Logs. We start from ConceptNet triples which contain the *MotivatedByGoal* relation such as [“wait tables”, “MotivatedByGoal”, “make money”]. The left concept **wait tables** represents a potential action to perform in order to attain the corresponding goal **make money** in the right concept. For each ConceptNet goal, we extracted a list of actions from ConceptNet. In case of the goal **make money**, the list of actions includes **wait tables**, **go to work** or **apply for a job**. To extend the knowledge base with human goals from Search Query Logs, we compiled a list of candidate human goals which were similar to the ConceptNet goal. We used a simple bag-of-words metric to identify similar goals: If the Search Query Log

Table 12: This table presents exemplary refinement candidates for selected ConceptNet goals. Frequencies from Search Query Logs are denoted in brackets and provide a first indicator towards quality of the candidates.

ConceptNet Goal	List of Refinements From Query Logs
now buy this car	buy new car (20), buy a used car (19), buy old cars (3), buy a cheap car (2), buying rental cars (2), buy electric car (2), buying wise car (1)
finding friends	find lost friend (11), to find old friends (4), find high school friends(1), find best friends (1), find elementary school friend (1), find free online friends (1), find past military friends (1)
writing a paper	write an argumentative paper (1), write an informative paper (1), write an autobiographical paper (1), write a narrative paper (1)
cutting your hair	cut short hair (3), cutting my own hair (1), cut black hair (1), cutting long hair (1)
feeding the baby	feeding a newborn baby (1)
find a partner	finding sexual partners (1)
train a dog	train an abused dog (1), train a deaf dog (1)
making coffee	make perfect coffee (1), make a flaming coffee (1)

goal contained all tokens of the ConceptNet goal, it was added to the candidate set of the corresponding ConceptNet goal. To give an example, **make money quickly** represent a candidate for the ConceptNet goal **make money**. This metric considers only refinements as candidates and consequently as potential new ConceptNet entries. The following example should clarify the procedure:

	ConceptNet Action	Relation Type	Similar Goals
ConceptNet Entry:	wait tables	MotivatedByGoal	make money
Potential New ConceptNet Entry:	wait tables	MotivatedByGoal	make money quickly

Table 13 shows potential new combinations of human goals from Search Query Logs and corresponding actions that were already contained in ConceptNet. The examples in Table 13 illustrate selected combinations that were rated positively by human judges. These combinations could be integrated into ConceptNet by using the *MotivatedByGoal* relation.

We conducted a human subject study to evaluate the compiled (“search query goal” / “ConceptNet action”) pairs which were annotated by two annotators. On the whole, 528 decisions had to be made. The human subjects were given a list of goals and corresponding actions. For every goal/action pair (G/A), they had to answer following question with yes or no: *Do you think that a person’s goal could be G when performing action A?* We introduced a softer variant of the known precision metric, i.e. a human goal from Search Query Logs was considered a potential goal if at least one ConceptNet action had been positively annotated. In our human subject study, we achieved an average soft-precision of 64%, meaning that 77 out of 120 goals from Search Query Logs were regarded reasonable goals for the given actions.

These findings suggest that human goals from Search Query Logs can con-

Table 13: This table shows selected combinations of human goals from Search Query Logs and corresponding actions from ConceptNet. These combinations were positively rated by human annotators and could be integrated into ConceptNet using the “MotivatedByGoal” relation.

List of ConceptNet Actions (Left Concept)	Semantic Relation	Human Goal from Search Query Logs (Right Concept)
wait tables, go to work, work the box office, serve customers, tell a story, get a contract, buy a house, apply for a job, pass a course	MotivatedByGoal	make some money quickly
meet interesting people, meet people	MotivatedByGoal	make new friends in your area
surf the net, surf the web, use a computer	MotivatedByGoal	find credit information
eat ice cream	MotivatedByGoal	ways to gain weight
go jogging, eat healthily, release your energy, go for a run, play sports, get exercise, get some physical activity, eat vegetables	MotivatedByGoal	lose maximum weight fast

tribute to complement ConceptNet, a commonsense knowledge base. In this section, we outlined how to generate promising ConceptNet node candidates and how to integrate them into ConceptNet.

## 6. Discussion

In this section, we will discuss threats to validity of our results (Yin, 2002). By examining our approach from a scientific perspective, we are capable of estimating the scientific value of our results, e.g. ensuring reproducibility.

### 6.1. Threats To Validity

*Construct validity:* The main construct we investigate in this research is the notion of *explicit goals contained in Search Query Logs*. While our definition intentionally gives some room for variability, our human subject study yields reasonable scores for inter-rater agreement  $\kappa$  (Cohen, 1960) and a reasonable distribution of human annotators’ judgements, which can be interpreted as preliminary empirical evidence for the validity of our construct.

*Internal validity:* The subjects involved in our human subject studies were graduate students enrolled at Austrian universities, who were not involved in the research of this paper. While all subjects were fluent in English, they did not share the same cultural context of the population that submitted those queries contained in the Search Query Logs (Austrian Graduate students vs. North-American web users). Although we could not find evidence of this problem in our human subject study, it could be the case that certain queries were mistakenly labeled as containing an explicit goal (such as “making the band”). Our bias towards longer queries ( $n > 2$ ) prevents us from studying a large part of the Search Query Logs ( $\sim 65\%$ ). Yet, the focus on longer queries was motivated by the observation that queries containing explicit goals mostly consist of more

than two tokens.

*External validity:* In our definition of “queries containing explicit human goals”, we refer to existing work on goal definitions from other research areas including goal-oriented requirements engineering (Regev & Wegmann, 2005; Liaskos et al., 2006) and computational linguistics (Tatu, 2005).

*Reliability:* In our experiments, we used existing toolkits such as WEKA (Witten & Frank, 2005) and NLTK<sup>9</sup> and established methods such as the Stanford part-of-speech tagger<sup>10</sup> and the Porter stemmer (Porter, 1997), so that reproducing our results should be possible.

## 7. Conclusions

Our work illustrates the potential of Search Query Logs to address challenges associated with acquiring knowledge about human goals from the web. Since Search Query Logs are a natural by-product of human activity on the web, they represent a largely untapped, renewable resource for the knowledge acquisition task. We present an automatic classification approach and demonstrate that human goals can be acquired from Search Query Logs with useful precision/recall values. Our results indicate that our presented automatic approach is a viable alternative to manual or semi-automatic approaches which are often costly and time-consuming. Our findings reveal that (i) human goals acquired from Search Query Logs in part represent commonsense goals which can be employed to refine and extend commonsense knowledge bases (Strohmaier & Kröll, 2009) and that (ii) they cover a vast range of topics and levels of granularity, which makes Search Query Logs a promising resource for addressing the goal coverage problem.

Our work is particularly relevant for knowledge engineers who seek to construct knowledge bases of human goals. In applications, goal knowledge can benefit various domains including text analysis (Kröll et al., 2010), human computer interaction (Smith & Lieberman, 2010) or planning (Tenorth et al., 2010). These applications would benefit from additional knowledge about human goals, e.g. recognizing a sequence of related human goals. Planning tasks involve the generation of action sequences that implement goals. Human goals from Search Query Logs can be employed to complement commonsense knowledge (see Section 5), which then could be integrated into artificial intelligence systems to support tasks such as planning. The potential of goal knowledge to inform human computer interaction is already being investigated. By equipping user interfaces with knowledge about human goals (Faaborg & Lieberman, 2006; Liu et al., 2002), a better understanding about users’ actions can be achieved contributing to advance the vision of more intelligent user interfaces.

---

<sup>9</sup><http://www.nltk.org/> accessed Feb 28th, 2011.

<sup>10</sup><http://nlp.stanford.edu/software/tagger.shtml> accessed Feb 28th, 2011.

### Acknowledgements

We would like to thank Microsoft Research for providing the search query log and Peter Prettenhofer for participating in completing this work. Furthermore, we would like to thank the reviewers for very constructive and detailed comments. This work is funded by the FWF Austrian Science Fund Grant P20269 TransAgere. The Know-Center is funded within the Austrian COMET Program under the auspices of the Austrian Ministry of Transport, Innovation and Technology, the Austrian Ministry of Economics and Labor and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency FFG.

### References

- Agichtein, E. & Gravano, L. (2000). Snowball: Extracting relations from large plain-text collections. In *Proceedings of the 5th ACM Conference on Digital Libraries (DL '00)*.
- Allan, J. & Raghavan, H. (2002). Using part-of-speech patterns to reduce query ambiguity. In *Proceedings of the 25th International Conference on Research and Development in Information Retrieval (SIGIR'02)*.
- Anderson, C. (2004). The long tail. *Wired Magazine*, 12(10).
- Anderson, L. W. & Krathwohl, D. A. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Blooms taxonomy of educational objectives*. New York, Longman.
- Baeza-Yates, R., Calderon-Benavides, L., & Gonzalez-Caro, C. (2006). The intention behind web queries. In *String Processing and Information Retrieval (SPIRE'06)*.
- Banko, M., Cafarella, M. J., Soderland, S., Broadhead, M., & Etzioni, O. (2007). Open information extraction from the web. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI'07)*.
- Banko, M. & Etzioni, O. (2007). Strategies for lifelong knowledge extraction from the web. In *Proceedings of the 4th International Conference on Knowledge Capture (K-CAP'07)*.
- Battelle, J. (2005). *The search: How Google and its rivals rewrote the rules of business and transformed our culture*. New York: Penguin Group.
- Beitzel, S., Jensen, E., Frieder, B., Lewis, D., Chowdhury, A., & Kolcz, A. (2005). Improving automatic query classification via semi-supervised learning. In *Proceedings of the 5th IEEE International Conference on Data Mining (ICDM'05)*.



- Brin, S. (1999). Extracting patterns and relations from the world wide web. In *Selected papers from the International Workshop on World Wide Web and Databases (WebDB '98)*.
- Broder, A. (2002). A taxonomy of web search. *SIGIR Forum*, 36(2).
- Carlson, A., Betteridge, J., Kisiel, B., Settles, B., Hruschka Jr., E., & Mitchell, T. (2010). Toward an architecture for never-ending language learning. In *Proceedings of the 27th National Conference on Artificial Intelligence (AAAI'10)*.
- Chulef, A. S., Read, S. J., & Walsh, D. A. (2001). A hierarchical taxonomy of human goals. *Motivation and Emotion*, 25(3).
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 1–37.
- Dai, H., Zhao, L., Nie, Z., Wen, J., Wang, L., & Li, Y. (2006). Detecting online commercial intention. In *Proceedings of the 15th International Conference on World Wide Web (WWW'06)*.
- Downey, D., Dumais, S., Liebling, D., & Horvitz, E. (2008). Understanding the relationship between searchers' queries and information goals. In *Proceedings of the International Conference on Information and Knowledge Management (CIKM'08)*.
- Dumais, S., Platt, J., Heckerman, D., & Sahami, M. (1998). Inductive learning algorithms and representations for text categorization. In *Proceedings of the International Conference on Information and Knowledge Management (CIKM'98)*.
- Eslick, I. (2006). *Searching for commonsense*. Master's thesis at the Massachusetts Institute of Technology.
- Etzioni, O., Banko, M., Soderland, S., & Weld, D. S. (2008). Open information extraction from the web. *Communications of the ACM*, 51(12), 68–74.
- Etzioni, O., Cafarella, M., Downey, D., Popescu, A., Shaked, T., Soderland, S., Weld, D. S., & Yates, A. (2005). Unsupervised named-entity extraction from the web: An experimental study. *Artificial Intelligence*, 165(1), 91–134.
- Faaborg, A. & Lieberman, H. (2006). A goal-oriented web browser. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI'06)*.
- Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.
- Friedman, J. (1997). On bias, variance, 0/1 loss, and the curse-of-dimensionality. *Data Mining and Knowledge Discovery*.

- Guo, Q. & Agichtein, E. (2010). Ready to buy or just browsing?: detecting web searcher goals from interaction data. In *Proceeding of the 33rd International ACM Conference on Research and Development in Information Retrieval (SIGIR '10)*.
- Havasi, C. (2007). Conceptnet 3: A flexible, multilingual semantic network for common sense knowledge. In *Proceedings of the 22nd Conference on Artificial Intelligence*.
- He, K., Chang, Y., & Lu, W. (2007). Improving identification of latent user goals through search-result snippet classification. In *Proceedings of the International Conference on Web Intelligence*.
- Hearst, M. (1992). Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th Conference on Computational Linguistics*.
- Hearst, M. (1998). Automated discovery of wordnet relations. In Fellbaum, C. (Ed.), *WordNet: An Electronic Lexical Database and Some of its Applications*. MIT Press.
- Hindle, D. (1990). Noun classification from predicate-argument structures. In *Proceedings of the 28th annual meeting on Association for Computational Linguistics (ACL'90)*.
- Hu, J., Wang, G., Lochovsky, F., Sun, J., & Chen, Z. (2009). Understanding user's query intent with wikipedia. In *Proceedings of the 18th International Conference on World Wide Web (WWW'09)*.
- Hui, B. & Yu, E. (2005). Extracting conceptual relationships from specialized documents. *Data And Knowledge Engineering*, 54(1), 29–55.
- Jansen, B. & Booth, D. (2010). Classifying web queries by topic and user intent. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems (CHI'10)*.
- Jansen, B., Booth, D., Park, D., Zhang, Y., Kathuria, A., & Bonner, P. (2009). To what degree can log data profile a web searcher? *Proceedings of the American Society for Information Science and Technology*.
- Jansen, B., Booth, D., & Smith, B. (2009). Using the taxonomy of cognitive learning to model online searching. *Information Processing and Management*, 45.
- Jansen, B., Booth, D., & Spink, A. (2008). Determining the informational, navigational, and transactional intent of web queries. *Information Processing and Management*, 44(3), 1251–1266.
- Jansen, B. & McNeese, M. D. (2005). Evaluating the effectiveness of and patterns of interactions with automated searching assistance: Research articles. *Proceedings of the American Society for Information Science and Technology*, 56, 1480–1503.

- Jansen, B., Spink, A., & Saracevic, T. (2000). Real life, real users, and real needs: A study and analysis of user queries on the web. *Information Processing and Management*, 36, 207–227.
- Jones, R. & Diaz, F. (2007). Temporal profiles of queries. *ACM Transactions on Information Systems*, 25.
- Kang, I. & Kim, G. (2003). Query type classification for web document retrieval. In *Proceedings of the 26th International Conference on Research and Development in Information Retrieval (SIGIR'03)*.
- Kim, J. & Moldovan, D. I. (1995). Acquisition of linguistic patterns for knowledge-based information extraction. *IEEE Transactions on Knowledge and Data Engineering*, 7(5), 713–724.
- Kirsh, D. (1990). When is information explicitly represented? *Information, Language and Cognition - The Vancouver Studies in Cognitive Science*, 340–365.
- Kröll, M., Körner, C., & Strohmaier, M. (2010). itag: Automatically annotating textual resources with human intentions. *Emerging Technologies in Web Intelligence*, 2(4), 333–342.
- Kröll, M. & Strohmaier, M. (2009). Analyzing human intentions in natural language text. In *Proceedings of the 5th International Conference on Knowledge Capture (K-CAP '09)*.
- Law, E., Mityagin, A., & Chickering, M. (2009). Intentions: A game for classifying search query intent. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems (CHI '09)*.
- Lee, U., Liu, Z., & Cho, J. (2005). Automatic identification of user goals in web search. In *Proceedings of the 14th International Conference on World Wide Web (WWW'05)*.
- Lenat, D. B. (1995). CyC: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11), 33–38.
- Levenshtein, V. I. (1966). Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *Soviet Physics Doklady*, 10.
- Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. University of Chicago Press.
- Li, X., Wang, Y., & Acero, A. (2008). Learning query intent from regularized click graphs. In *Proceedings of the 31st International ACM Conference on Research and Development in Information Retrieval (SIGIR '08)*, (pp. 339–346)., New York, NY, USA. ACM.
- Li, Y., Zheng, Z., & Dai, H. (2005). KDD cup-2005 report: Facing a great challenge. *ACM SIGKDD Explorations Newsletter*, 7(2), 91–99.

- Liaskos, S., Lapouchnian, A., Yu, Y., Yu, E., & Mylopoulos, J. (2006). On goal-based variability acquisition and analysis. In *Proceedings of the 14th IEEE International Requirements Engineering Conference (RE'06)*.
- Lieberman, H. (2008). Usable AI requires commonsense knowledge. In *Workshops and Courses: Usable Artificial Intelligence held in conjunction CHI'08: Proceedings of the Conference on Human Factors in Computing Systems*.
- Lieberman, H., Liu, H., Singh, P., & Barry, B. (2004). Beating common sense into interactive applications. *AI Magazine*, 25(4).
- Lieberman, H., Smith, D., & Teeters, A. (2007). Common consensus: A web-based game for collecting commonsense goals. In *Proceedings of the Workshop on Common Sense and Intelligent User Interfaces held in conjunction with the 2007 International Conference on Intelligent User Interfaces*.
- Liu, H., Lieberman, H., & Selker, T. (2002). Goose: A goal-oriented search engine with commonsense. In *Proceedings of the 2nd International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems*.
- Liu, H. & Singh, P. (2004). ConceptNet - A practical commonsense reasoning tool-kit. *BT Technology Journal*, 22(4), 211–226.
- Murray, G. & Teevan, J. (2007). WWW workshop report: Query log analysis - social and technological challenges. *ACM SIGIR Forum*, 41(2).
- Pantel, P. & Pennacchiotti, M. (2008). Automatically harvesting and ontologizing semantic relations. In *Proceedings of the 2008 Conference on Ontology Learning and Population: Bridging the Gap between Text and Knowledge*, (pp. 171–195)., Amsterdam, The Netherlands, The Netherlands. IOS Press.
- Pasca, M., Van Durme, B., & Garera, N. (2007). The role of documents vs. queries in extracting class attributes from text. In *Proceedings of the International Conference on Information and Knowledge Management (CIKM '07)*.
- Pass, G., Chowdhury, A., & Torgeson, C. (2006). A picture of search. In *Proceedings of the 1st International Conference on Scalable Information Systems*.
- Passonneau, R. J. & Litman, D. (1993). Intention-based segmentation: Human reliability and correlation with linguistic cues. In *Proceedings of the 31st annual meeting on Association for Computational Linguistics (ACL'93)*.
- Porter, M. F. (1997). An algorithm for suffix stripping. In *Readings in Information Retrieval*, (pp. 313–316)., San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.
- Potts, C., Takahashi, K., & Anton, A. (1994). Inquiry-based requirements analysis. *Software, IEEE*, 11(2), 21–32.

- Regev, G. & Wegmann, A. (2005). Where do goals come from: The underlying principles of goal-oriented requirements engineering. In *Proceedings of the 13th International Conference on Requirements Engineering (RE'05)*.
- Riloff, E. (1996). Automatically generating extraction patterns from untagged text. In *Proceedings of the 13th National Conference on Artificial Intelligence (AAAI'96)*.
- Rose, D. E. & Levinson, D. (2004). Understanding user goals in web search. In *Proceedings of the 13th International Conference on the World Wide Web (WWW'04)*.
- Schank, R. & Abelson, R. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Lawrence Erlbaum Associates.
- Singh, P., Lin, T., Mueller, E., Lim, G., Perkins, T., & Zhu, W. (2002). Open mind common sense: Knowledge acquisition from the general public. In *Proceedings of the 1st International Conference on Ontologies, Databases, and Applications of Semantics for Large Scale Information Systems*.
- Smith, D. (2007). *EventMinder: A personal calendar assistant that understands events*. Master's thesis at the Massachusetts Institute of Technology.
- Smith, D. & Lieberman, H. (2010). The why UI: Using goal networks to improve user interfaces. In *Proceedings of the 14th International Conference on Intelligent User Interfaces (IUI'10)*.
- Soderland, S., Fisher, D., Aseltine, J., & Lehnert, W. (1995). CRYSTAL: Inducing a conceptual dictionary. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI'95)*.
- Strohmaier, M. & Kröll, M. (2009). Studying databases of intentions: Do search query logs capture knowledge about common human goals? In *Proceedings of the 5th International Conference on Knowledge Capture (K-CAP'09)*.
- Strohmaier, M., Kröll, M., & Körner, C. (2009). Intentional query suggestion: Making user goals more explicit during search. In *Proceedings of the 2009 Workshop on Web Search Click Data (WSCD'09)*.
- Strohmaier, M., Lux, M., Granitzer, M., Scheir, P., Liaskos, S., & Yu, E. (2007). How do users express goals on the web? - An exploration of intentional structures in web search. In *International Workshop on Collaborative Knowledge Management for Web Information Systems*.
- Strohmaier, M., Prettenhofer, P., & Kröll, M. (2008). Acquiring explicit user goals from search query logs. In *International Workshop on Agents and Data Mining Interaction ADMI'08, in conjunction with WI '08*.
- Suchanek, F., Kasneci, G., & Weikum, G. (2007). Yago: A core of semantic knowledge. In *Proceedings of the 16th International Conference on World Wide Web (WWW'07)*.

- Tatu, M. (2005). Automatic discovery of intentions in text and its application to question answering. In *Proceedings of the 43rd annual meeting of the Association for Computational Linguistics (ACL'05)*.
- Teevan, J., Dumais, S. T., & Liebling, D. J. (2008). To personalize or not to personalize: Modeling queries with variation in user intent. In *Proceedings of the 31st annual International ACM Conference on Research and Development in Information Retrieval (SIGIR '08)*.
- Tenorth, M., Nyga, D., & Beetz, M. (2010). Understanding and executing instructions for everyday manipulation tasks from the world wide web. *IEEE International Conference on Robotics and Automation*.
- Vapnik, V. (1998). *Statistical Learning Theory*. New York, NY, USA Wiley-Interscience.
- Witten, I. H. & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann Series in Data Management Systems, Morgan Kaufmann, 2nd ed.
- Yi, X., Raghavan, H., & Leggetter, C. (2009). Discovering users' specific geo intention in web search. In *Proceedings of the 18th International Conference on World Wide Web (WWW '09)*.
- Yin, R. (2002). *Case study research: Design and methods (Applied Social Research Methods)*. SAGE Publications.