My research focuses on the computational foundations of intelligent behavior. We develop theories and systems pertaining to intelligent behavior using a unified methodology – at the heart of which is the idea that learning has a central role in intelligence.

We attempt to understand the role of learning in supporting intelligent inference and use this understanding to develop systems that learn and make intelligent inferences in complex domains. Such systems would acquire the bulk of their knowledge from raw, real world data, and behave robustly when presented with new, previously unseen, situations. These systems can be studied at various levels of abstraction and from various viewpoints. We have concentrated on developing the theoretical basis within which to address some of the obstacles and on developing an experimental paradigm so that realistic experiments (in terms of scale and resources) can be performed to validate the theoretical basis. The main area of concentration in this realm has been that of natural language understanding and intelligent access to textual information.

My work spans several aspects of this problem – from foundational questions in learning, knowledge representation and reasoning, to experimental paradigms and large scale system development – and draws on methods from theoretical computer science, probability and statistics, artificial intelligence, linguistics and experimental computer science. We are driven both by longer term goals to understand and develop capabilities for natural language comprehension, and by challenging shorter term applications in the area of information extraction and knowledge access.

In terms of “traditional” research areas my work falls mostly into Natural Language Processing and Machine Learning, but also to Learning Theory, Knowledge Representation and Reasoning. In the last few years we have been developing the following interrelated lines of work, all pertain to the idea that learning is of prime importance in performing knowledge intensive inferences and use the natural language domain as the main application area.

- Learning in Natural Language: We have addressed several key problems in natural language from a unified point of view and developed both a theoretical understanding of several fundamental issues and an experimental paradigm that builds on this view. In particular, we have moved from studying stand alone natural language predication problems to higher level problems such as semantic parsing and textual entailment. These efforts have also driven the development of several successful systems and natural language processing tools.

- Learning and Inference: Continuing my long standing line of work on integrating theories of learning and inference we have developed an integrated learning and inference approach in the context of natural language processing, addressing natural language problems that are global in that multiple interrelated components at several levels of abstraction affect the inference.

- Intelligent Information Access: Developing better methods to access knowledge represented in text is one of the forces driving the integrated learning and inference paradigm above; we have addressed multiple fundamental information access problems from this perspective.

- Knowledge Representation and Inference: We study intermediate knowledge representations that facilitate learning and inference in complex domains. We have also developed novel inference formulations and algorithms for complex probabilistic and relational representations including, for the first time, a relational inference algorithm for probabilistic representations.
• Learning Theory: Much of the research described above builds on our work in learning theory and drives the foundational learning theory questions we address. Over the last few years we have addressed learning problems with constrained output, ranking problems and kernels over structures.

Perhaps the most significant aspect of this research program is in helping to place theoretical work in learning and inference in the context of realistic inference problems and thus contributing to a shift in focus in the natural language research community, a move to more advanced learning paradigms, and to work on higher level problems. Along with developing the theoretical foundations we have contributed to the development of a practical approach to building large-scale learning-centered intelligent systems and NLP tools and have developed advanced tools that are being used by many researchers and in industry.

1 Learning in Natural Language

We describe below some of the background, progress and impact of our work on Machine Learning in Natural Language. More details on some of the research directions that we have pursued are presented below in Sections 2, 3 and 4.

Linear Classifiers and Discriminatory Learning: Early empirical work in natural language processing was influenced by the success of statistical speech recognition and was dominated by relatively simple statistical methods. Many of the early works, from statistical part-of-speech tagging [11, 12] to the noisy channel model for machine translation, have roots in work conducted in the speech field. Most of the early works can be viewed as based on generative probability models, which provide a principled way to study statistical classification. In these models, it is common to assume a generative model for the data, estimate its most likely parameters from training data and then use Bayes rule to obtain a classifier for this model. Naturally, estimating the most likely parameters involves making simplifying assumptions about the generating model.

My earlier work in this area has contributed to developing better understanding of the relations between probabilistic models of classification and discriminative models and has had a significant effect on work in natural language processing [49, 50, 20].

In [49, 50, 51, 52] we have shown that the decision surface of many probabilistic classifiers (and other classifiers used in earlier work in natural language) is linear over some feature space. We have used these observations to (1) develop a learning theoretical explanation for the success of some probabilistic methods despite the clear failure of their assumptions and (2) suggested that one should keep using the same linear representation, but develop other methods of parameter estimation, driven directly by the eventual goal: to support better predictions.

These observations, which are now considered “common knowledge” in machine learning, have influenced the SNoW learning architecture which we have developed [49, 9] – consisting of enhanced, regularized versions of the perceptron and winnow algorithms and of naive Bayes – which has been downloaded by thousands of researchers around the world. More importantly, this understanding has contributed to a vast use of discriminative approaches and a range of linear classifiers such as Boosting, SVMs, Winnow and Perceptron, all successfully applied to a broad range of natural language problems.

The significance of these results goes beyond explaining the generalization and robustness properties of widely used methods. Rather, they provide insight into possible extensions of these methods (1) to learn from more structured, knowledge intensive observations, as part of a learning centered approach to higher level natural language inferences and (2) to learn more structured output, where
multiple output variables represent interdependent problem components. We describe below, and in more details in Sec. 2, our work in the latter direction and then, in Sec. 4 our work on the former.

**Learning with Structured Output:** The emphasis on discriminative methods applies not only to simple classification problems but also to machine learning work on more complex structured models. In [50] we showed that predicting the most likely state in a HMM (or other graphical probabilistic models) has a linear decision surface over features that correspond to transition probabilities and state-observations probabilities. Using dynamic programming this observation immediately leads to a discriminative algorithm for training models with structured output. Indeed, this work has motivated our work on learning with structured output (see below) [39, 55, 56, 57, 43, 44], and has influenced other models that incorporate dependencies among the variables into the learning process, and directly induce estimators to optimize an appropriate performance measure [13]. Our recent work in this line of research is described in Sections 2 and 3 below.

**Software Tools and NLP Solutions:** Along with developing theoretical understanding, models and algorithms for problems in natural language processing, we have developed a number of mature tools that were made available to the research community and have been downloaded and used by thousands of researchers, and are being used in computational linguistics classes and in industry. In addition to our basic machine learning package, SNoW, and a feature extraction language, FEX, we have made available a collection of state-of-the-art NLP tools. In addition to some important general purpose pre-processing tools such as a sentence segmentation tool, we have made available one of the best part-of-speech taggers and one of the best shallow parsers available. Other tools include a name-entity recognizer, along with an applet for annotation, a question analysis and classification tool and others. A large number of on-line demos, both for the above mentioned packages and others were also made available and are being frequently used, including in NLP and Computational Linguistics classes. These include a context sensitive speller and the best semantic parser available (see details below; the latter is a real-time version of the approach that won the first place in the shared task competition in the Conference of Learning in Natural Language in June 2005.). Tools, packages and data are all available from http://L2R.cs.uiuc.edu/~cogcomp/.

## 2 Learning and Inference

Our earlier work on learning and inference has focused on the Learning to Reason (L2R) framework – an integrated theory of learning, knowledge representation and reasoning within a unified framework which uses learning to facilitate high level reasoning tasks [25, 46, 26, 47, 22]. We view our current work on learning and inference as a concrete instantiation of the earlier framework in the sense that L2R served to (1) highlight the importance of the knowledge representation as a facilitator of learning and inference processes and the lack of commitment to a specific, comprehensible one and (2) to show that it is possible to support global inference even when learning the representation isn’t exact, by exploiting global properties of the task.

In the last few years, we have concentrated on Learning and Inference over structured and constrained output. Many problems in natural language processing involve assigning values to sets of variables where a complex and expressive structure can influence, or even dictate, what assignments are possible. Assigning part-of-speech tags to the words of a sentence and many phrase segmentation and parsing problems, are key examples of such tasks. Traditionally, solutions to these problems were generative, as represented by HMM models for sequence learning problems and syntactic parsing. More recently, when discriminative methods for classification became popular, structured problems were addressed this way too, by decoupling learning from the task of maintaining structured output. Only after estimators are learned for each local output variable are
they used to produce global output consistent with the structural constraints.

Our first work in this direction, within a framework we called *Inference with classifiers* [?] was [39]. We proposed and studied three models for inference with classifiers under sequential constraints (This work was a runner-up for the best paper award is NIPS, given to papers co-authored by a student). In this work we started to move away from the study of single classifiers and studied ways to use classifiers in downstream inference processes. We studied the problem of combining the outcomes of several different classifiers in a way that provides a coherent inference that satisfies some constraints. Beyond developing a Markovian approach that extends standard HMMs to allow the use of a rich observation structure and of general classifiers to model state-observation dependencies, the most promising approach we developed is an extended constraint satisfaction based formalism for combining different classifiers into a single coherent decision that respects both the classifiers’ predictions and multiple constraints that their outcomes ought to satisfy. This approach led to the development of one of the best shallow parsers available [39, 40, 10, 34].

However, in many cases, the interactions between the variables of interest (e.g., variables representing the semantic role of phrases in a sentence, subject, object, temporal adjunct, etc.) cannot be expressed in a sequential manner, and thus inference to determine them cannot be done using dynamic programming techniques. This is the case when one wants to express long term dependencies among variables, number constraints (e.g., this verb cannot take more than three arguments), etc.

Over the last few years we have developed a fairly general setting to address these problems [55, 56, 57, 42, 43, 44, 41]. We defined the problem in terms of a collection of discrete random variables representing binary relations and their arguments; we seek an optimal assignment to the variables in the presence of the constraints on the binary relations between variables and the relation types. We have formalized these problems as constrained optimization problems over variables that represent the outcome of learned classifiers, and have shown how to model *Inference as Linear Programming*, or Integer Linear Programming problems, yielding exact and efficient solutions.

We studied this approach theoretically and empirically and addressed multiple questions such as: (1) efficiency and modularity of the approach and issues related to integrality of solutions [56]; (2) comparing different training paradigms for learning structured output [42] e.g., we compared paradigms that decouple training from inference with paradigms that couple training and inference tightly, at the cost of losing efficiency and modularity [30, 13]; (3) incorporating both statistical and declarative constraints within our framework [57]; and (4) the view of the framework as a generalization of sequential approaches (e.g., Viterbi), in the sense that when the constraints are less expressive, the inference reduces to simpler computation with polynomial complexity guarantees [57].

Most of this research was done in the context of developing an approach to the task of Semantic Role Labeling - a shallow semantic analysis of sentences at the level of “who did what to whom, when, where and how” [43, 42, 41]. The system we developed for this task based on the aforementioned approach was the top ranked system in the shared task competition run by the Conference on Natural Language Learning in June 2005, out of 19 international groups participating. (An earlier version developed in 2004 won the second place.)

Our work on Inference as Constraint Optimization, and even the specific (Integer) Linear Programming formulation, already has had significant impact in natural language research, from information extraction work to works in natural language generation [38].
3 Intelligent Access to Information

Most of today’s knowledge is available in a textual form. This information is typically accessed in one of two ways: key-word search (via search engines) or “semantic queries” presented to a database, in the rare case that the information is available in the form of a database with a known schema. The goal of our research in this area is to apply progress we make in our work on learning in natural language in order to bridging this gap, and allow access to free form text as if it was in a database with a known schema. To accomplish this goal we attempt to recognize some level of the semantics of the free form text and use it to develop better ways to access this information.

In the last few years we have concentrated on two problems in this direction. The first, Recognizing Entities and Relations is the problem of identifying entities in text (e.g., identifying phrases that represent names of people, locations, organizations, etc.) as well as the relations between them – e.g., this sentence indicates that A is the AUTHOR of B or that C LIVES in D. The second problem, Identifying and Tracking Entities Across Documents builds on a relatively robust identification of mentions of entities in documents, and focuses on the ability to identify the entity itself, within single documents and across documents. Namely, we would like to be able to determine that the strings JFK, President Kennedy, John Kennedy, Kennedy (and other variations) all refer to the same person and, at the same time, determine based on the context that, sometimes, John Kennedy (the baseball player) refers to a different person. A third effort under the general Information Access line of research, on Question Answering, will be discussed in Sec. 4 below.

Our work on Recognizing Entities and Relations is done within the Integer Linear Programming inference approach described above. The key technical problem we wanted to address when studying this information extraction approach is that of moving beyond the pipeline architecture. A pipeline approach is the typical strategy employed in solving complex natural language problems – separating a task into several stages and solving them sequentially, where the features in stage \(i\) typically reflect a commitment to predictions made in stage \(i - 1\). For example, a named entity recognizer may be trained in advance, using some training data, and then given as a black box to a relation classifier, to be used as a feature extraction tool. This is often done at multiple levels, starting with tokenization, segmentation, part-of-speech tagging, etc. Clearly, this strategy disregards interactions across layers and propagation of error. Our approach, therefore, aims at developing a way in which multiple stages in this pipeline can interact, reaching a simultaneous final global decision on all the variables of interest. Specifically, we studied the problem of simultaneously recognizing named entities and relations between them, and have shown that the inference approach allows us to support bidirectional interaction between these stages, via the integer linear programming paradigm we developed [55, 56].

The second focus of our work in intelligent information access in the last few years was on the problem of Identifying and Tracking Entities Across Documents. A given entity – representing a person, a location or an organization – may be mentioned in text in multiple, ambiguous ways. Supporting concept-based (rather than “string-based”) access to information requires resolving conceptual ambiguity and, in particular, identifying whether different mentions of real world entities, within and across documents, actually represent the same concept.

We developed several machine learning based approaches to this problem [31, 32, 33, 36] that differ in the amount of supervision they require, in efficiency of training the models, and (as it turns out) also in the robustness to parameter initialization and tuning. Along with addressing this problem we have also begun to develop a new approach to training clustering functions, as described below. Our first approach to the problem is a global generative model [32], at the heart of which is a view on how documents are generated and how names (of different entity types) are “sprinkled” into
them. In its most general form, our model assumes: (1) a joint distribution over entities (e.g., a document that mentions “President Kennedy” is more likely to mention “Oswald” or “White House” than “Roger Clemens”), (2) an “author” model, that assumes that at least one mention of an entity in a document is easily identifiable, and then generates other mentions via (3) an appearance model, governing how mentions are transformed from the “representative” mention.

We then developed a way to train this model in a discriminative manner [31] requiring the training of a pairwise local classifier in a supervised way, to determine whether two given mentions represent the same real world entity. This is followed, potentially, by a global clustering algorithm that uses the classifier as its similarity metric. Following lessons from these two works we have developed a new approach that is based on a new view of clustering [36]. Clustering is an optimization procedure that partitions a set of elements to optimize some criteria, based on a fixed distance metric defined between the elements. The inherent noise in the data used in the entity identification problem has motivated us to develop a new view of clustering, in which clustering is viewed as a learning, rather than only an optimization task; we proposed a way to train a distance metric that is appropriate for the chosen clustering algorithm in the context of the given task.

Our work on this problem is summarized in an invited paper to the AI magazine [33] and has resulted in significant on-going collaboration with database researchers, with the hope of incorporating these ideas in the context of semantic integration of databases.

An additional effort with the Intelligent Information Access, on Question Answering, has focused both on our machine learning work on analysis and classification of questions [35, 37, 36], and on textual entailment [7, 8], and is discussed below.

4 Knowledge Representation and Inference

Our work in this area over the last few years has three main directions. First, we studied knowledge representations and relational learning over structured domains [29, 14, 54, 15, 17, 16, 57]. Progress in this direction has led to the development of a description-logic inspired representation that has been the key representation we have been using for the development of the second direction – our work on an inference models for textual entailment [7, 8]. Finally, we have made a significant progress on the problem of inference with probabilistic first-order representations; we developed the first lifted probabilistic inference algorithm [18] – one that does not need to resort to propositionalization.

Intermediate Representations that facilitate learning: Learning becomes easy once the correct input representation has been chosen, for example, one that produces linearly separable point sets. Our work here focuses on automatically generating intermediate representations to aid supervised learning algorithms. Most of the machine learning work in natural language (as in other domains like machine vision) is done over structured input, where attributes, as well as relations among them, need to be encoded as input to the learning algorithm. Most of the computation in machine learning is thus spent at the stage of feature extraction - both off line, in order to determine what types of features are important, and on-line, as the feature extraction process is often more time consuming than training and evaluation.

We developed a representation language for learning over structured domains within a proposition-alization framework that decouples feature construction and model construction. We formalized this in two different and complementary ways, that address several aspects of the problem. First, we developed and studied a flexible knowledge representation for structured data, with an associated language that provides the syntax and a well defined equivalent semantics for expressing complex structured data succinctly. Second, we used this language to automate the process of feature con-
struction by expressing types of objects which are instantiated in the ground data, allowing us to
determine the level at which learning is done. Finally, this process of re-representation of domain
elements allows general purpose learning schemes, such as feature efficient linear algorithms and
probabilistic representations, to be defined over the resulting space, yielding efficient and expressive
learning of relational functions over a structured domain using efficient propositional means.

The Feature EXtraction language that we designed, FEX, has been made available to the research
community, and has been downloaded by a large number of researchers.

Our formalism uses description logics and concept graphs in the service of learning relational models
using efficient propositional learning algorithms. We introduced Feature Description Logic (FDL) - a
relational (frame based) language that supports efficient inference, along with a generation
function that uses inference with descriptions in the FDL to produce features suitable for use by
learning algorithms. These are used within a learning framework that is shown to efficiently and
accurately learn relational representations in terms of the FDL descriptions. The paradigm was
designed and shown effective in support of learning in domains that are relational but where the
amount of data and size of representation learned are very large. This paradigm provides a natural
solution to the problem of learning and representing relational data; it extends and unifies several
lines of works in KRR and Machine Learning in ways that provide hope for a coherent usage of
learning and reasoning methods in large scale intelligent inference.

Complementing our work on relational intermediate representations described above is the work on
kernels for structured data [17]. This work addresses also the tradeoff between the computational
efficiency with which these kernels can be computed and the generalization ability of the resulting
classifier. This study complements our theoretical study of these issues, in the context of studying
structured kernels for on-line learning algorithms [27, 28].

Indicative of the usefulness of our approach – originally developed in the context of natural language
problems – is its success in facilitating learning in visual recognition problems [53, 5, 1], done as
part of our attempt to study learning in multi-modal environments. The knowledge representation
and learning approach are identical in the natural language and visual recognition domains, with
the only difference being the sensory level (sensors, formally defined objects within the language
as a mapping from observations to “formulas” [54]). Our visual recognition work on relational
representations within a discriminatory learning framework [5, 1] (also known as a constellation
model), has already had a significant impact on this research area.

Textual Entailment: Semantic entailment is the problem of determining if the meaning of a given
sentence entails that of another. For example, determine if the sentence: WalMart defended itself
in court today against claims that its female employees were kept out of management
because they are women entails the sentence: WalMart was sued for sexual discrimination. This
is a fundamental problem in natural language understanding that provides a broad framework for
studying language variability and has a large number of applications. Over the last year we have
started to develop a principled approach to this problem that builds on inducing representations of
text snippets into a hierarchical knowledge representation along with a sound optimization-based
inferential mechanism that makes use of it to decide semantic entailment.

[7, 8] present a principled computational approach to semantic entailment in natural language,
addressing some of the key problems encountered in traditional approaches, such as knowledge ac-
quisition and brittleness. The solution uses a hierarchical knowledge representation language into
which we induce appropriate representations of the given text and required background knowledge.
The other main element is a sound inferential mechanism that makes use of the induced repre-
sentation to determine an extended notion of subsumption, using an optimization approach that
supports abstracting over language variability and representation inaccuracies.
The knowledge representation we study in the context of this problem is an Extended Feature Description Logic, an extension of the description logic based representation we alluded to earlier as the knowledge representation we developed for structured domains. Our view of inference as an optimization procedure builds on the constrained optimization formalisms described in Sec. 2.

**Lifted First Order Probabilistic Inference:** We study inference algorithms and representations for expressive, first order, probabilistic reasoning. This work continues our early, high impact work on the complexity of probabilistic inference [45, 48] and attempts to provide solutions to some of the key barriers to the use of expressive probabilistic representations in large scale problems.

Most probabilistic inference algorithms are specified and processed on a propositional level. In the last decade, many proposals for more expressive, first order probabilistic representations were made, but none came with an algorithmic solution to the inference problem. In the inference stage they all still operate on a propositional representation level, resulting in a huge blow-up in the representation size in a way that eliminates any possible gain from the compact first order representation. This makes inference impossible for realistic size problems. Our work is the first exact inference algorithm (following ideas from [Poole, 2003] that applied for a special case) that operates directly on a first-order level, and that can be applied to any first-order model (specified in a language that generalizes undirected graphical models). This recent work [18] promises to revolutionize inference with large relational probabilistic representations.

5 Learning Theory

A lot of the work described above - where we study machine learning and inference methods in the context of making progress in natural language - builds on our work in Learning Theory. Over the last few years we have concentrated on four main problems in this area, some of which have already had direct impact on machine learning for natural language and other areas.

**Kernels over structured data** [27, 28, 17]. Specifically, we looked at efficiency versus convergence properties of kernels for on-line learning algorithms.

One method of increasing the expressiveness of learned hypotheses is to expand the feature set to include conjunctions of basic features. This can be done explicitly or, where possible, by using a kernel function. Focusing on the well known Perceptron and Winnow algorithms, we demonstrated a tradeoff between the computational efficiency with which the algorithm can be run over the expanded feature space and the generalization ability of the corresponding learning algorithm [27, 28]. In [17] we focused on structured domains, designed a family of kernels for these, and studied the question of when, as a function of the domain size and the number of examples, is it better to use structured kernels, vs. an explicit generation of features.

**Multi-class Classification:** [23, 24]. We showed that the dominant ways machine learning practitioners study multi-class classification are not expressive, but that a slight modification of a commonly used algorithm is, provably, the right way to approach multi-class classification. In doing that, we developed the constraint classification framework which captures many flavors of multiclass classification including winner-take-all multiclass classification, multilabel classification and category ranking. We presented a meta-algorithm for learning in this framework that learns via a single linear classifier in high dimension, and presented distribution independent and margin-based generalization bounds for this algorithm, as well as empirical and theoretical evidence showing the advantage of constraint classification over existing methods of multiclass classification.

**Margin Distribution, Generalization Bounds and Learning Algorithms** [19, 21].

The study of generalization abilities of learning algorithms and their dependence on sample com-
plexity is one of the fundamental research efforts in learning theory. Understanding the inherent difficulty of learning problems allows one to evaluate the possibility of learning in certain situations and to estimate the degree of confidence in the predictions made and is thus crucial in understanding and analyzing learning algorithms.

Understanding generalization is even more important when learning in very high dimensional spaces, as in many natural language and computer vision applications. Specifically, can one count on the behavior of a $10^6$ dimensional classifier that is trained on a few examples, or even a few thousands examples? In [19, 21] we studied generalization properties of linear learning algorithms and developed a data dependent approach that is used to derive generalization bounds that depend on the margin distribution. Our method uses random projection techniques to allow the use of existing VC dimension bounds in the effective, lower, dimension of the data, resulting in bounds that are tighter than existing bounds. We showed that these results can be used as a model selection criteria and presented the Margin Distribution Optimization (MDO) learning algorithm, that directly optimizes this complexity measure.

**Learning Ranking Functions and Generalization Bounds for the Area Under the ROC Curve** [6, 4, 3, 2]. In a series of recent works we have studied ranking problems and, specifically, generalization properties of learning ranking functions and their appropriate complexity measures. In many learning problems, the goal is not simply to classify objects into one of a fixed number of classes; instead, a ranking of objects is desired. This is the case, for example, in information retrieval problems, where one is interested in retrieving documents from some database that are ‘relevant’ to a given query or topic. In such problems, one wants to return to the user a list of documents that contains more relevant documents at the top and less relevant documents at the bottom; in other words, one wants a ranking of the documents such that relevant documents are ranked higher than irrelevant documents.

The problem can be shown to be different than the traditionally studied classification problem. We concentrated on the question of what is a good ranking function? and showed that this question is also different than the analogous well studied question for classification, in the sense that a good classification function may not always translate into a good ranking function. We went on to study generalization properties of the area under the ROC curve (AUC), a quantity that has been advocated as an evaluation criterion for the (bipartite) ranking problem.

## 6 Summary

While our research falls into several traditional research areas – Machine Learning, Natural Learning Processing, Knowledge Representation and Reasoning and Learning Theory – it constitutes a coherent line of research, advancing the study of intelligent behavior, using a unified methodology. At the heart of our approach is the idea that learning has a principle role in supporting intelligent behavior, and we develop theories that advance our understanding, and systems that implement these ideas and investigate them in the context of realistic and important real world applications, mostly in the context of advancing natural language understanding.

Our progress over the last few years offers some hope that machines can be instilled with complex knowledge in areas where efforts to achieve this by programming alone have not succeeded. Along with developing the theoretical foundations our work has contributed to placing theoretical work in learning in the context of realistic inference problems and to the development of a practical approach to building large-scale intelligent systems, as in some of the solutions developed to key problems in natural language.
References


