An F-Measure for Context-based Information Retrieval

Michael Kandefer and Stuart Shapiro
Department of Computer Science and Engineering, Center for Cognitive Science,
Center for Multisource Information Fusion
University at Buffalo, Buffalo, NY 14260
{mwk3,shapiro}@cse.buffalo.edu

Abstract

Computationally expensive processes, such as deductive reasoners, can suffer performance issues when they operate over large-scale data sets. The optimal procedure would allow reasoners to only operate on that information that is relevant. Procedures that approach such an ideal are necessary to accomplish the goal of commonsense reasoning, which is to endow an agent with enough background knowledge to behave intelligently. Despite the presence of some procedures for accomplishing this task one question remains unanswered: How does one measure the performance of procedures that bring relevant information to bear in KR systems?

This paper answers this question by introducing two methods for measuring the performance of *context-based information retrieval* processes in the domain of KR systems. Both methods produce an *f-measure* as a result. These methods are evaluated with examples and discussion in order to determine which is more effective. Uses of these measures are also discussed.

Introduction

Computationally expensive processes, such as deductive reasoners, can suffer performance issues when they operate over large-scale data sets. The optimal procedure would allow such processes to operate with only the information that is relevant to the current task. Bringing relevant information to bear has numerous applications in context-aware agents/devices (Arritt & Turner 2003; Dey 2001; Bradley & Dunlop 2005; Dourish 2004; Kurz, Popescu, & Gallacher 2004).[1] In KR systems, reasoning is probably most hampered in large-scale knowledge bases due to complicated procedures, like building and maintaining search trees resulting from knowledge base queries. Such concerns with large-scale knowledge bases have been discussed previously (Subramanian, Greiner, & Pearl 1997) and various solutions have been offered (Haarslev & Moller 2001; Levy, Fikes, & Sagiv 1997; Lenat 1998; 1995). Due to these concerns, a method for including a minimal set of background knowledge for the current task is necessary if we are to accomplish the goals of commonsense reasoning, which is to endow an agent with all the commonsense knowledge necessary to exhibit intelligent behavior. Methods for solving this problem have been proposed or implemented in the past (Anderson 2007), and others are capable of being implemented in KR systems (Arritt & Turner 2003). However, one question remains unanswered: How does one measure the performance of procedures that bring relevant information to bear in KR?

This paper answers this question by discussing two methods for measuring the performance of *context-based information retrieval* (CBIR) processes in the domain of KR systems:

- Relevance Theoretic Measure, and
- Distance from the Optimal

The first is based on a method for determining the relevance of a subset of an agent's knowledge base given contextual information. Sperber and William (1995) initially proposed the *relevance-theoretic* approach for use in modeling communication, but believe it has uses in other cognitive processes. Harter (1992) agrees with this notion, but claims that the approach is also useful for determining relevance in information retrieval (IR) testing. Borlund and Ingwersern (1997) agree, but limit this type of testing to a particular type of relevance called "situated relevance". The *distance from the optimal* is our own method.

There are several uses for measurement methods like the above. The foremost is comparing the results of various CBIR procedures. Cohen (1995) has noted that we often do not know if a program has worked well, or poorly. Such evaluations often deal with speed and space considerations, but in CBIR procedures we are also interested in measuring the utility of the results. The measures above are one way of accomplishing this.

Other than comparing CBIR procedures, many CBIR procedures, such as spreading activation (Howes 2007; Crestani 1997; Loftus 1975) and context diagnosis (Arritt & Turner 2003), operate by utilizing various parameters that can be set to influence their performance. These parameters are given arbitrary values and then tested to find suitable levels. The above measurements schemes can aid in the process, and potentially make it automatic.

Though the two measurement methods can be used for

[1] Here "context" is not the knowledge representation and reasoning (KR) sense of the term, but defined as "the structured set of variable, external constraints to some (*real* or *artificial*) cognitive process that influences the behavior of that process in the agent(s) under consideration" (Kandefer & Shapiro 2008).

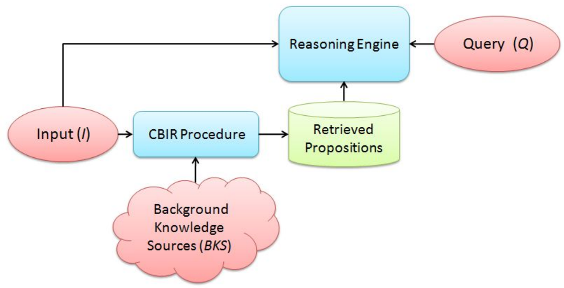

Figure 1: Context-Based Information Retrieval Process

the above tasks, our interest in this paper is in determining which makes a more effective tool for evaluating CBIR results. In order to accomplish this we will calculate the *f-measure* values of these methods when applied to example CBIR results. A *f-measure* is the standard measure for evaluating IR results.

Context-Based Information Retrieval

CBIR is an independent, preprocessing step that occurs before reasoning. A general CBIR procedure operates by examining an input, typically sensory. It uses that input to constrain the knowledge that is available to the reasoner. This process is depicted in Fig. 1.

The CBIR procedure receives input (*I*), which contains the *contextual constraints* and other information about the situation; and the background knowledge (*BKS*) containing any knowledge that will be evaluated by the CBIR procedure. With this the CBIR procedure produces a subset of the background knowledge, called *retrieved propositions*, for use by a reasoning engine, that can then be queried (*Q*). These queries could also be expected goals an embodied agent should be capable of achieving in context.

In "The Handle Problem" domain (Miller & Morgenstern 2006), which is the problem of inferring whether or not an object can be used as a handle through a description of its properties and relationships with other objects, an example of such input would be spatial information about some objects that might be door handles. Such information could contain a unique identifier for an unidentified object, the object's shape (e.g., conical, rectangular, etc.), and feature information (e.g., whether the object is inverted, or blunt). An example of background knowledge in such a domain would include various assertions about using objects with certain properties as handles. Though the CBIR procedure ultimately produces information for consumption by the reasoner, the tasks of the reasoner can also influence what information should be retrieved. This is apparent in goal seeking situations, such as question answering. As such, some CBIR procedures take into account the goals of the agent or the state of a problem they are solving and provide these as input.

The most important aspect of the process requires that the CBIR procedure output any *retrieved propositions*, which will be used by the reasoning engine. As such, the knowledge provided as relevant by the CBIR procedure will always be a subset of the *BKS*. This information is selected by the CBIR procedure through an algorithm that examines the *BKS* and *I*. This algorithm varies between CBIR procedures, but it should be noted that most do not examine the entirety of the *BKS*, but only an initial subset determined by *I*. The *retrieved propositions* determines a successful CBIR procedure, and what we will evaluate.

Measuring Results

As previously mentioned the output of the CBIR procedure is a subset of the *BKS*, called the *retrieved propositions*, and a means of establishing successful results is required. In information retrieval (IR) the accepted practice for evaluating such results is to calculate an *f-measure*. An *f-measure* score is between 0.0 and 1.0, with 0.0 indicating the poorest result and 1.0 a perfect retrieval. An *f-measure* identifies situations where IR results contain unnecessary information, called *precision*, and where the results do not contain enough information, called *recall*. In order to calculate an *f-measure* (Fig. 2) for CBIR results the *retrieved propositions* and another set of propositions, called the *relevant propositions*, are necessary. Below two methods for acquiring a set of *relevant propositions* and using them for evaluating the *retrieved propositions* from a CBIR process are discussed. In both of these methods the process of generating the set of *relevant propositions* can be accomplished any time prior to the calculation of the *f-measure* for the CBIR results.

Relevancy Theory

Relevancy Theory (Sperber & Wilson 1995) is a model developed by Wilson and Sperber in the field of pragmatics that is used for explaining the cognitive process listeners undertake as they approach an understanding of a speaker's utterance. A system implementing this model is said to be using a *relevance-theoretic* method. The *relevance-theoretic* approach is not limited to establishing the relevance of an utterance, but also of observable phenomena, memories, and current thoughts. The process of determining relevancy relies on a principle that states something is relevant to a cognitive agent, if the agent can utilize it to draw conclusions that matter to it. When such conclusions are reached this is said to be a *positive cognitive effect*.

In relevance theory these *positive cognitive effects* are utilized to *measure* the degree of relevancy of a particular input, where an input could be any of the cognitive artifacts discussed above. However, most of the discussion by Wilber and Sperber has focused on the communication aspects of relevancy, and determining when an utterance is relevant to the current working memory contents of an agent. The working memory of the agent is represented as a set of propositions, which are a subset of the contents of the agent's BKS. These assumptions are used by William and Sperber to define *positive cognitive effects*. We take that definition, but modify it slightly so it can be used for determining the set of relevant propositions in an agent's BKS, rather than an utterance the agent encounters. We

Recall (r)	Precision (p)	F-measure (F)
$r = \frac{\vert\{relevant\ propositions\}\cap\{retrieved\ propositions\}\vert}{\vert\{relevant\ propositions\}\vert}$	$p = \frac{\vert\{relevant\ propositions\}\cap\{retrieved\ propositions\}\vert}{\vert\{retrieved\ propositions\}\vert}$	$F(r,p) = \frac{2rp}{r+p}$

Figure 2: Formulas for computing the *f-measure* (van Rijsbergen 1979)

Entire Knowledge Base
$A1 : \forall(x,y)(Blunt(x) \wedge Conical(x) \wedge Drawer(y) \wedge ConnectedByTip(x,y) \rightarrow Handle(x))$.
$A2 : \forall(x)(Handle(x) \rightarrow CanBePulled(x))$.
$A3 : Blunt(h1)$.
$A4 : Conical(h1)$.
$A5 : \forall(x,y)(Rope(x) \wedge Light(y) \wedge Connected(x,y) \rightarrow CanBePulled(x)$
$A6 : \forall(x,y)(Blunt(x) \wedge Conical(y) \wedge ConnectedByBase(x,y) \rightarrow \neg Handle(x)$
$A7 : \forall(x)(Drawer(x) \rightarrow ContainsItems(x))$.

Figure 3: The entire knowledge base

define a *positive cognitive effect* as follows (italicized words will be discussed below):

Given I and Q, as sets of propositions, and *BKS*, then if there is a proposition *p* that is an element of BKS, but not an element of $\{I \cup Q\}$, then p is a *positive cognitive effect* if either:

1. $\neg p \in \{I \cup Q\}$,
2. *p* helps *strengthens* some q that is an element of $\{I \cup Q\}$, or
3. *p* contributes to a *contextual implication*, which is defined as the condition where:
   (a) $\{\{I \cup Q\} \cup BKS\}$ *non-trivially* derives using p some proposition q, and
   (b) $\{I \cup Q\}$ alone does not *non-trivially* derive q, and
   (c) BKS alone does not *non-trivially* derive q

In case (1) a comparison between the $\{I \cup Q\}$ and BKS is made that determines if any of the propositions contradict one another. Each proposition in BKS that does is considered a *positive cognitive effect*. Case (2) involves a notion of strengthening that can occur when two sets of propositions are compared. The strengthening of proposition q in $\{I \cup Q\}$ occurs when: (1) $\{\{I \cup Q\} \cup BKS\}$ *non-trivially* derives q, or (2) BKS *non-trivially* derives q, which was derived in $\{I \cup Q\}$ already. Any propositions that are members of BKS and that are involved in such derivations are counted as *positive cognitive effects*. Case (3) establishes as *positive cognitive effects* those propositions in BKS that are involved in a *non-trivial* derivation using propositions from both $\{I \cup Q\}$ and BKS, which can not be done by $\{I \cup Q\}$ or BKS independently.

Of the three cases two rely on a notion of a *non-trivial* derivation. A formalization of this concept is not trivial, not provided by William and Sperber, and beyond the scope of this paper. For the sake of simplicity we will consider any proposition involved in a *modus ponens* rule of inference to be *non-trivial* in our examples.

With the above method for establishing *positive cognitive effects* the *relevance-theoretic* approach can be used for measuring the relevancy of the set of *retrieved propositions* from a CBIR procedure. To accomplish this the above method is used to find all the *positive cognitive effects* in BKS. This resulting proposition set is taken as the *relevant propositions*. With the *retrieved propositions* and *relevant propositions* available the *recall*, *precision*, and *f-measure* can be calculated for each CBIR output using the formulas in Fig. 2.

To illustrate, assume we have the KB depicted in Fig. 3 as the BKS (created by us from propositions that might be useful for solving "The Door Handle Problem") and three retrieved proposition sets: *Usable Conical Drawer Handles*, *Conical Drawer Handles*, and *Misc. Handles and Drawers*. Assume also that these were output as relevant from three different CBIR procedures, and that they have the propositional content depicted in Fig. 4.

Suppose the following proposition set is a combination of the expected input and query, $\{I \cup Q\}$, to the agent in context: $\{Drawer(d1) \wedge ConnectedByTip(h1, d1) \wedge CanBePulled(h1)\}$. With this the *relevance theoretic* approach determines that $\{A1, A2, A3, A4, A7\}$ are the *relevant propositions* of the background knowledge sources as they are involved in part of *contextual implications* that result in the derivation of *Handle(h1)*, *CanBePulled(h1)* , and *ContainsItems(d1)*. With this the *f-measure* can be calculated for each CBIR *retrieved propositions* set. This is done using the cardinality of the *retrieved propositions* (Ret.), the cardinality of the *relevant propositions* (Rel.), and the cardinality of their intersection (Int.). (Fig. 5). For example, the *Usable Conical Drawer Handles* has retrieved four propositions that are all in the relevant proposition set. As such, the intersection between he two is also four and it receives a *precision* of 1.0 (4/4). However, there are five relevant propositions, as such the *recall* is 0.8 (4/5).

Given the results of the *f-measure* calculation, the retrieved proposition sets that is most relevant would be *Usable Conical Drawer Handles*. As such the relevancy method that retrieved that proposition set would be deemed better at the CBIR procedure than the other two.

Usable Conical Drawer Handle
$A1 : \forall(x, y)(Blunt(x) \wedge Conical(x) \wedge Drawer(y) \wedge ConnectedByTip(x, y) \rightarrow Handle(x)).$ $A2 : \forall(x)(Handle(x) \rightarrow CanBePulled(x)).$ $A3 : Blunt(h1).$ $A4 : Conical(h1).$
Conical Drawer Handles
$A1 : \forall(x, y)(Blunt(x) \wedge Conical(x) \wedge Drawer(y) \wedge ConnectedByTip(x, y) \rightarrow Handle(x)).$ $A2 : \forall(x)(Handle(x) \rightarrow CanBePulled(x)).$ $A3 : Blunt(h1).$ $A4 : Conical(h1).$ $A6 : \forall(x, y)(Blunt(x) \wedge Conical(y) \wedge ConnectedByBase(x, y) \rightarrow \neg Handle(x)$
Misc. Handles and Drawers
$A2 : \forall(x)(Handle(x) \rightarrow CanBePulled(x)).$ $A3 : Blunt(h1)$ $A4 : Conical(h1).$ $A5 : \forall(x, y)(Rope(x) \wedge Light(y) \wedge Connected(x, y) \rightarrow CanBePulled(x)$ $A7 : \forall(x)(Drawer(x) \rightarrow ContainsItems(x)).$

Figure 4: Three different outputs from different context-sensitive retrieval operations.

Retrieved Proposition Set	Rel.	Ret.	Int.	Recall	Precision	F-Measure
Usable Conical Drawer Handles	5	4	4	0.8	1.0	0.889
Conical Drawer Handles	5	5	4	0.8	0.8	0.8
Misc. Handles and Drawers	5	5	4	0.8	0.8	0.8

Figure 5: The results of calculating the *f-measure* using a relevance theoretic approach.

Distance from the Optimal

Distance from the optimal is a method of testing that examines the input to a system and creates the optimal results on which to compare a system's future performance. In the CBIR model, if given a reasoning query (i.e., a particular reasoning task given to the reasoner) *Q*, the input propositions *I*, the contents of the background knowledge sources *BKS*, and a reasoner, that is capable of keeping track of the origin sets,[2] or equivalent, then the optimal solution for the original query can be calculated. This is accomplished by the following algorithm:

1. Given some query proposition Q that the reasoner is asked to derive, the entire knowledge base BKS that the CBIR procedure would access, and an input I that the CBIR procedure would use to produce its output.
2. Load the BKS into the reasoner.
3. Add I to the BKS.
4. Query the reasoner on Q.
5. Examine the origin set for Q, OS_Q, defined as:[3]

$$OS_Q = \{A - I | A \subset \{BKS \cup I\} \wedge \\ A \vdash Q \wedge \\ \neg\exists(A')(A' \subsetneq A \wedge A' \vdash Q)\}.^4$$

6. Select the sets in OS_Q that have the minimal cardinality. This new set of origin sets will be denoted with $min(OS_Q)$.[5]

After this process is complete we have those origin sets that derive Q, and that also contain the minimal number of propositions needed to do so. Since these propositions are necessary for reasoning to the desired conclusion and minimal, we shall consider any origin set in the set of minimal solutions an *optimal solution*. With the possible optimal solutions in hand, we can measure the results of a CBIR procedure against each optimal solution and compute a *f-measure* for the results.

The presence of multiple optimal solutions poses some problems for computing the *f-measure*. To handle this *recall*, *precision*, and *f-measure* must be calculated treating each optimal solution as the *relevant propositions* and then comparing it to the CBIR output, or the *retrieved propositions*. The highest *f-measure* is chosen as the result. The reason for choosing the highest is that the CBIR output might share few propositions with some of the optimal solutions, but still match one of them precisely. In such a scenario the CBIR

[2] An origin set for a proposition is the set of propositions used in the derivation of that proposition. Origin sets originate from *relevance logic* proof theory (Shapiro 1992).

[3] $A \vdash B$ indicates that a proposition B can be derived from the set of propositions A.

[4] *I* is removed since in a CBIR procedure it is automatically provided to the reasoner and it should not impact retrieval scores.

[5] This step is performed as there can be multiple reasoning "paths" to Q in a BKS that use different proposition sets.

Retrieved Proposition Set	Rel.	Ret.	Int.	Recall	Precision	F-Measure
Usable Conical Drawer Handles	4	4	4	1.0	1.0	1.0
Conical Drawer Handles	4	5	4	1.0	0.8	0.889
Misc. Handles and Drawers	4	5	3	0.75	0.6	0.667

Figure 6: The results of calculating the *f-measure* using the distance from the optimal approach.

output is at least capable of generating one of the perfect solutions. Formulas for *recall*, *precision*, and the *f-measure* are the same as those used in the *relevance-theoretic* approach (Fig. 2).

To illustrate how this measure can be used for evaluating the results of CBIR procedures consider an example using the the knowledge base depicted in Fig. 3 as the BKS parameter in the above algorithm. Let I be the proposition: $ConnectedByTip(h1, d1) \wedge Drawer(d1)$ and Q the query $CanBePulled(h1)?$.[6] After execution of the query we receive one origin set for Q: $\{A1, A2, A3, A4\}$, and since it is the only one, it is inserted into $min(OS_Q)$. With these values calculated we can now compare the optimal solution against the CBIR procedure outputs. We will again use the ones discussed in Fig. 4. The results are depicted in Fig. 6.

Since *Usable Conical Drawer Handles* is the actual optimal solution OS_Q it gets a perfect *f-measure* of 1.0. The *Conical Drawer Handles* receives the next highest as it had a perfect *recall*, but contain one extraneous proposition affecting its *precision*. The last retrieved proposition set, *Rope Handles*, was penalized heavily as it did not retrieve all of the relevant propositions (*recall*) and contained numerous propositions that weren't part of the relevant proposition set (*precision*).

Evaluation

Though both methods measure the same unit (i.e., propositions) and rely on rules of inference to ultimately create a score for the results, they differ in their generation of the *relevant propositions*, and thus, the *f-measure*. The greatest difference is that the *relevance-theoretic* uses the input to find all possible propositions that trigger *positive cognitive effects*, while the distance from the optimal only looks for the minimal set needed. This can result in needed discrimination when measuring CBIR results. In the example this is illustrated when the *relevance-theoretic* approach provided the same score for *Conical Drawer Handles* and *Misc. Handles and Drawers*, while the *distance from the optimal* provides a useful distinction.

The *relevance-theoretic* approach also values higher those CBIR outputs that contain multiple solutions to the same problem, since all propositions involved in those solutions would cause positive cognitive effects, despite the fact that only one solution is needed. This ultimately causes more reasoning and more computation time, which is what we would like a CBIR procedure to avoid. The *distance from the optimal* method values CBIR procedures that produce close to optimal solutions, and ones with multiple solutions would be considered as having extraneous propositions.

Finally, one important difference between the two methods is that the *relevance-theoretic* approach requires a formalization of a *non-trivial* deduction. This is not an easy task, as it involves determining which rules of inference and which combinations of them are trivial.

Measurement Requirements

For the two methods presented for measuring CBIR in knowledge representation and reasoning (KR) a KR system is needed to perform the actual measurements. This system need not be the same as the one in the process diagram. It requires the ability to:

- **Store and reason over a large number of propositions.** The CBIR methods are designed to retrieve relevant information from larger knowledge bases and present them to a reasoner to limit processing. While this design is done to eliminate the need for a KR system to have all of the background information available to it, measuring the success of CBIR processes does need a KR system to reason over the entire knowledge base every time a *relevant proposition* set needs to be generated. While the term "large" is vague a knowledge base with approximately 100,000 propositions causes problems for some reasoning tasks. Speed of reasoning is not required for the measurements.
- **Perform forward and backward chaining.** Both measurements rely on forward chaining, and the *distance from the optimal* relies on backward chaining to generate the set of *relevant propositions*, which are used in measuring the results of the CBIR procedures.
- **Detect non-trivial deductions.** The *relevance-theoretic* approach requires that the KR system recognize *non-trivial* derivations in order to prevent the mislabeling of trivial propositions in that derivation as relevant. These *non-trivial* derivations are needed to populate the set of *relevant propositions*.
- **Compute and store the origin sets of derived propositions.** The *distance from the optimal* measurement requires that the KR system keep track of the minimal number of propositions required to derive another proposition (i.e., the origin set) in order to create the set of *relevant propositions*.

Conclusions and Future Work

The *relevance-theoretic* approach for determining the relevant propositions in a knowledge base, an approach that has been proposed as a useful method for determining relevancy

[6] This differs slightly from the previous example since the *relevance theoretic* approach does not take into account how the retrieved propositions will be used (e.g., expected queries, agent goals).

in information retrieval, was found to be less successful than the *distance from the optimal* method for measuring the results of CBIR procedures. This was mostly because the *relevant theoretic* approach finds all solutions to a problem and marks all propositions involved in those solutions as relevant, while the *distance from the optimal* finds the minimal number. Some theoretical issues also hinder the *relevance-theoretic* approach. Its reliance on *trivial implications* is one such issue, as they are difficult to properly formalize. This formalization step is necessary prior to development of a tool that can use the *relevance-theoretic* approach for measuring the results of CBIR procedures. The long-term goals of commonsense reasoning will require methods for retrieving a subset of an agent's background knowledge based on context. CBIR procedures address this issue, and choosing method for measuring their performance will be required.

Apart from these findings, a theoretical discussion and an example was used to compare the two measures. Examples like this serve as a useful precursor to the development of test cases for evaluating the measures against each other. In the future we will explore such test cases. In doing so, a formalization of *non-trivial* deductions will also be produced.

References

Anderson, J. R. 2007. Human associative memory. In *How Can the Human Mind Occur in the Physical Universe?* NY, New York: Oxford University Press. 91–134.

Arritt, R., and Turner, R. 2003. Situation assessment for autonomous underwater vehicles using a priori contextual knowledge. In *Proceedings of the Thirteenth International Symposium on Unmanned Untethered Submersible Technology (UUST)*.

Borlund, P., and Ingwersern, P. 1997. The development of a method for the evaluation of interactive information retrieval systems. *The Journal of Documentation* 53(3).

Bradley, N. A., and Dunlop, M. D. 2005. Toward a multidisciplinary model of context to support context-aware computing. *Human-Computer Interaction* 20:403–446.

Cohen, P. R. 1995. Empirical research. In *Empirical Methods for Artificial Intelligence*. The MIT Press,Cambridge, MA.

Crestani, F. 1997. Application of spreading activation techniques in information retrieval. *Artif. Intell. Rev.* 11(6):453–482.

Dey, A. K. 2001. Understanding and using context. *Personal and Ubiquitous Computing* 5(1):4–7.

Dourish, P. 2004. What we talk about when we talk about context. *Personal Ubiquitous Computing* 8(1):19–30.

Haarslev, V., and Moller, R. 2001. High performance reasoning with very large knowledge bases: A practical case study. In *International Joint Conference on Artificial Intelligence*.

Harter, S. P. 1992. Psychological relevance and information science. *Journal of the American Society for Information Science* 43(9).

Howes, M. B. 2007. *Human Memory*. Sage Publications, Inc. chapter Long-term Memory – Ongoing Research, 131–160.

Kandefer, M., and Shapiro, S. C. 2008. A categorization of contextual constraints. In Samsonovich, A., ed., *Biologically Inspired Cognitive Architectures, Technical Report FS–08–04*, 88–93. AAAI Press, Menlo Park, CA.

Kurz, B.; Popescu, I.; and Gallacher, S. 2004. FACADE - a framework for context-aware content adaptation and delivery. In *Proceedings of Communication Networks and Services Research*, 46–55.

Lenat, D. B. 1995. Cyc: A large-scale investment in knowledge infrastructure. *Communications of the ACM* 38(11).

Lenat, D. 1998. The Dimensions of Context Space. Technical Report, Cycorp.

Levy, A. Y.; Fikes, R. E.; and Sagiv, Y. 1997. Speeding up inferences using relevance reasoning: a formalism and algoirthms. *Artificial Intelligence* 97(1–2):pg. 83–136.

Loftus, A. C. E. 1975. A spreading activation theory of semantic processing. *Psychological Review* 82(6):407–428.

Miller, R., and Morgenstern, L. 2006. Common sense problem page. `http://www-formal.stanford.edu/leora/commonsense/`.

Shapiro, S. C. 1992. Relevance Logic in Computer Science. In Anderson, A. R.; Nuel D. Belnap, J.; and Dunn, J. M., eds., *Entailment*, volume II. Princeton, NJ: Princeton University Press. pg. 553–563.

Sperber, D., and Wilson, D. 1995. *Relevance: Communication and Cognition*. Blackwell Publishing, Malden, MA.

Subramanian, D.; Greiner, R.; and Pearl, J. 1997. The relevance of relevance. *Artificial Intelligence* 97(1–2):pg. 1–5.

van Rijsbergen, C. 1979. *Information Retrieval, Second Edition*. Butterworths.