A case-based representation of legal text
for conceptual retrieval

Judith P. Dick and Graeme Hirst
Faculty of Library and Information Science (Dick) and
Department of Computer Science (Hirst)
University of Toronto
Toronto, Ontario
Canada M5S 1A1

dick@cs.toronto.edu and *gh@cs.toronto.edu*

Abstract

Our research constitutes an attempt to move from
state-of-the art IR toward conceptual retrieval in
the domain of case law. A knowledge representa-
tion has been developed using Toulmin's argument
structure, Sowa's conceptual graphs, and Somers's
case grammar. A frame matcher is used to demon-
strate retrieval using questions derived from law
cases following those in the knowledge base. The
goal throughout has been to provide the legal
researcher with access to cases, and help him or
her to develop an argument. In our opinion, intel-
ligent retrieval involves making it easier for the
user to find his or her own way conceptually.

1. Introduction

Natural language processing techniques have been
used in our recent research (Dick 1991) to demon-
strate how an information retrieval system can be
designed to help people find better information in a
more natural, intuitive way. The user should be
able to locate information, the concepts that
interest him, as well as the documents. Moreover,
he or she should be able to develop his or her ideas
as he or she searches with a minimum of interfer-
ence from the system. In our opinion, an intelli-
gent system would free its user to think while
seeking information, unconstrained by its logic,
linguistic and mechanical limitations. A concep-
tual retrieval system ought to be designed to suit a
real person with changing ideas, and would adapt
itself to his or her thinking as it progresses.

Present day information retrieval systems
(IRS) are designed with the intention of satisfying
the needs of the greatest number of users as deter-
mined by statistical surveys of users' behavior in
using systems. As a result, they attempt to satisfy
the need of a prototypical user, the average user,
not the individual user, in the most efficient way
possible. In order to accomplish this they depend
on keyword representations and Boolean logic in
order to access databases consisting of large

volumes of text. The users of these systems rou-
tinely get in response to their queries lists of docu-
ments about the subjects named in their requests.
The documents are selected through the use of
index terms, nominalizations, used in combination
to match terms from the documents in the data-
base. It is expected that the documents so selected
will probably be relevant to the query.

We intend to show that it is possible to
retrieve not only documents about named subjects,
but information about unnamed ideas and
knowledge implicit in the text if a more expressive
representation than keywords is used for text.
Retrieving conceptual information requires that the
meaning of the text be represented, not just the
names of the subjects it describes. Conceptual
retrieval is often confused with topical approach to
information finding, such as one might encounter
in using the subject heading approach to a library
catalog. What we mean by conceptual information
is not accessible by means of a topical index, it is
the underlying meaning of the text. In order to
access concepts, it is necessary to represent the
semantic content of the text; naming the subjects
the documents are about is not adequate.

In our research, a knowledge representation
(kr) has been developed, which focuses on seman-
tic analysis of contract law cases. The intended
users of our proposed system were legal research-
ers who search cases for authority to support a
point of view that will be argued in a case at hand.
Since the goal of case law retrieval is the develop-
ment of a new argument to suit a current need, the
highest level of analysis was the dissection of the
principal argument in each contract case. Stephen
Toulmin's 'good reasons' model (1958) was used
to structure the arguments. Within the argument
schemata, there are frames designed to accommo-
date a case-based representation of the text. John
Sowa's conceptual graphs (1984), and Harold
Somers's linguistic cases (1987), have been used
in constructing the kr. We have added a lexicon of
legal concepts, a catalog of conceptual relations
and rules for semantic selection.

The kr is the heart of the research. However, we have demonstrated its retrieval capability by describing the use of a suitable structure matcher, LOG+, to show how natural language queries can be answered by matching frames. Realistic questions were derived from the fact situations of real contract cases that later followed the cases in the knowledge base (kb).

2. Background

The literature indicates a continuum of development from document retrieval to full text IR. It starts with the keyword-Boolean systems, and ends with the situation that Meadow predicted for 2001, information scientists anticipating "a world of holistic recording in which the intellectual effort of deciding what is worth seeing comes after the recording" (1979, p. 218). Between those two extremes there are a number of other positions, among them, text passage retrieval (O'Connor, 1980), compromises between controlled vocabulary and kr (Karlgren 1977), and conceptual retrieval (Schank et al. 1981). There are indications in the literature of a growing interest in looking beyond the traditional IRS model.[1]

In order to achieve intelligent text retrieval, it is essential to move away from keyword indexes and toward more expressive representations of text based on linguistic concepts. The key to finding the 'middle-ground' between IR and NLP is, in our opinion, the development of a viable representation for meaning. The representation must be unambiguous, but coarse enough to highlight the informational content of the text rather than focus on the literary expression. Furthermore, we wanted it to be robust and potentially suitable for the handling of large quantities of text. We contend that the concepts to be represented must be derived from the text if retrieval is to be meaningful. If pre-selected terms were to be related to the text, the resulting system would not satisfy. Our encoding was done manually because the focus of the research was on the retrieval target rather than the process. We wanted to show that conceptual retrieval can be accomplished. The conceptual graphs notation we used for our representation is already supported by a growing body of software. We anticipate that in due course such a representation as this one will be automatically derivable from text. Similarly, the users' questions will be automatically encoded for searching. In this

research, a coarse-grained representation has been developed using caseframes that describe objects derived from the text itself.

3. The domain

Law cases have been chosen for our demonstration because the need for conceptual retrieval is obvious in this domain and pressing. Retrieving of cases has always been very difficult. The legal researcher first determines what the issues are in the problem at hand, then proceeds to isolate the important facts and focus on legal concepts that are significant. At this stage the argument that will ultimately result from his or her work is embryonic. It will develop during the course of the search for further information. The researcher needs to make associations among selected legal concepts and to navigate among various legal concepts with their related facts in order to investigate the issues in any given legal problem and to construct the necessary argument. Keyword-Boolean systems make it difficult to satisfy those needs. Both factual and topical indexes to law cases are exceptionally good. However, we have known the limitations of indexing for a very long time. In 1897, Oliver Wendell Holmes wrote the following anecdote:

> There is a story of a Vermont justice of the peace before whom a suit was brought by one farmer against another for breaking a churn. The justice took time to consider, and then said that he could had looked through the statutes and could find nothing about churns, and gave judgment for the defendant. The same state of mind is shown in all our common digests and textbooks. Applications of rudimentary rules of contract or tort are tucked away under the head of Railroads or Telegraphs or go to swell treatises on historical subdivisions, such as Shipping or Equity, or are gathered under the arbitrary title which is thought likely to appeal to the practical mind, such as Mercantile law. (Holmes 1897, p. 59.)

Clearly, information entailing the concept behind the incident could not have been found using a keyword-Boolean retrieval system. Moreover, Holmes discussed how unsatisfactory the use of subject headings was for locating conceptual information.

Since textual analysis in this domain is especially challenging, we contend that if conceptual retrieval can be done with law cases, it is possible with any other kind of text. Cases are particularly difficult to analyze as each is unique. Patterns of literary similarity are not common. The reasoning is diffuse, dense, and original. Although reasons

[1] See especially the Fall 1989 issue of *International journal of intelligent systems* (vol. 4, no. 3), a special issue concerning, "Knowledge-based techniques for information retrieval".

for judgement are formally written, the vocabulary of the law is derived from everyday language. When interpreting the language, it is exacting to distinguish the legal from the common meanings as both are often in use. In this way, the capability of our representation to distinguish between uses is strongly tested. Furthermore, the cases are replete with abstract terminology. Attaching meaning to simple concrete nominalisations such as 'tax' or 'budget' does not compare in difficulty with the semantic analysis of abstract legal concepts. Among the expressions included in the cases are the formidable concepts of 'intention to contract' and the 'foreseeability of consequences'.

The kb consists of decisions taken from a contracts casebook (Milner 1985). It was important that the group of cases share enough common concepts to make demonstrations of reasonably complex retrieval problems possible. The cases range from 1605 to 1942, and come from American, British, and Canadian jurisdictions. Some are simple, some complex. They provide a wide variation of rhetoric in both the reasoning and stylistic senses.

In the lexicon, definitions of legal concepts have been taken from legal dictionaries, treatises and other authoritative sources. The lexicon adds to the real world knowledge of the system. As well, each legal concept has explicit relations to information in the cases in which it is used. It functions as the core of a cluster of related concepts. The legal concepts develop as knowledge accumulates. They do not have single dimension, rigid definitions, but may present meaning from numerous perspectives. The legal concepts are open textured and will never be exhaustively defined.

In addition to the case representations and the lexicon of legal concepts, we have constructed semantic selection rules. They are derived from the case representations and restrict pattern matches with the concepts in questions. The function of the selection rules is to discriminate among the concepts in the kb and so choose those matching the meaning of the question.

The precise object of conceptual retrieval is to access information about what a concept, or combination of concepts, actually entails. The kr makes it possible to disambiguate meanings. Retrieval works by pattern matching, rather than sorting or categorization as do document retrieval systems. Because of the precision of the semantic representation it is possible in our system, unlike keyword systems, to distinguish among the concepts entailed by frequently used terms such as

'court', 'proceeding', and 'trial'. Partial descriptions of concepts, or fact situations can be used in retrieval. It is possible to retrieve information related to difficult abstract concepts.

4. The knowledge representation

The principal argument, the *ratio decidendi*, of each case was analyzed according to Toulmin's argument model (1958). The model was used in structuring the discourse. The argument schema consists of claim, grounds, reasons, backing, modal qualifiers and rebuttal.

Sowa's conceptual graphs (cgs) are used as notation for the representations and they provide the logical base of the system (Sowa 1984). We have used cgs for a number of practical reasons and a number of qualitative ones. A fully developed notation is available now. Cgs have a mnemonic aspect and are relatively easy for the uninitiated to read; their use makes bridging the gap between IR and AI audiences easier. Furthermore, there is an established user community and we believe that employing cgs brings us closer to the construction of an interpreter, because of current software development.

The expressiveness of the cgs makes them attractive for NLP. Furthermore, Sowa has not tied the graphs to a particular linguistic theory, but instead has demonstrated their versatility for semantic representation. Of particular interest in the context of our research, he accommodated the use of deep cases, which combine syntactic and semantic features and lend themselves to use with first-order logic (FOL).[2]

Case theory was introduced in 1968 by Fillmore. As Hirst explains

> In its most basic form, case theory views a sentence as an assertion whose predicate is denoted by the verb of the sentence and whose arguments are denoted by the noun phrases. (1987, p. 7)

No single case grammar has emerged as clearly superior. Nor has a consensus been achieved on a list of fundamental cases. The grid is designed to be used in computational analysis. As Somers put it, "Case is altogether a question of making significant generalizations" (1987, p. 119). He attempted to solve three case problems, the degree of specificity, the assignment of arbitrary names

---

[2] In this paper, we unavoidably use the word 'case' in two completely different senses: the legal sense of a particular trial and judgement, and the linguistic sense of an underlying semantic structure. It should be clear from the context which sense is being used when — even if both occur within a single sentence.

for exceptions and finally, role duality. He was able to answer the most ringing criticisms of case theory. The case grid appears in Table 1 at the end of this paper.

The row designations are characteristic grammatical relations. The columnar designations are semantic realizations. Each cell specifies the nature of the syntactic-semantic link expressed by that case. Each row and each column is to be read as a parameter of cases.

Within the framework of Sowa's near-FOL notation, and the valency-based case system of Somers, it is clear that a verb-centered approach to text analysis resulted in a powerful representation suitable for conceptual retrieval with the use of generalized.

5. Examples of the knowledge representation

There is at the highest level of organization, a case identification frame for each case in the kb. It contains essential information about the trial and the bibliographical information about the case. The actual kr of each case is structured according to the argument schema described above. The first is the case of *Weeks* v. *Tybald* (1605 Noy 11; 74 E.R. 982). The facts are straightforward:

> "In this case it would appear that the plaintiff or his father was told by the defendant, whose daughter the plaintiff later married, that he would give 100 pounds to him that should marry his daughter with his consent.' Held, for defendant. 'It is not averred nor declared to whom the words were spoken, and it is not reasonable that the defendant should be bound by such general words spoken to excite suitors.' "

The promise made by D[3] Weeks is shown in figure 1. As an introduction to the representation the excerpt may be read as follows. The promise, numbered '#W1' contains the terms as stated in the case. The concept [TERM] is used to keep the terms within a single context. In the antecedent of the conditional statement, the two propositions [MARRY] and [CONSENT_TO] are conjoined as they appear consecutively. They say that if some unknown man *x marries the daughter of the Tybald family, and if D Tybald consents to the event [MARRY:#W1], then that same D Tybald will give the man *x money in the amount of £100. The conceptual relation (ACTS) designates the use of the case active source, indicating the initiator and agent of the action. The daughter in the marry

---

[3]'D' will be used for defendant, and 'P' for plaintiff.

```
[PROMISE-n: #W1]—
  (CONT)—[TERM:
    if [MARRY: #W1]—
      (ACTS)—[MAN: *x]
      (ACTL)—[DAUGHTER: Tybald].
    [CONSENT_TO-v: #W1]—
      (ACTS)—[D: Tybald]
      (OBJL)—[MARRY: #W1],
    then [GIVE: #W1]—
      (ACTS)—[D: Tybald]
      (DATPOSSG)—[MAN: *x]
      (DATPOSSL)—[MONEY: @£100],].
```

Figure 1. The promise in *Weeks* v. *Tybald.*

---

event has is the value of the slot (ACTL) active local. It indicates that she is not the object of the marriage, but the co-agent. The two agents share equally in their part in the event—the difference between the cases is syntactic. The (ACTS) is the subject. The concept [CONSENT_TO-v: #W1] includes a verb and preposition, a particle construction. The preposition carries an essential part of the meaning of the verb phrase. The (OBJL) or objective local slot indicates that its filler, the value [MARRY: #W1] is an entity undergoing the process of [CONSENT_TO]. It is a passive role rather than active in this context. In the consequent of the condition, the event [GIVE] has three slots with the cases named active source (ACTS), dative possessive goal (DATPOSSG), and dative possessive local (DATPOSSL). (ACTS) is the agentive case, like those seen above. However, the (DATPOSS_) cases have another dimension. They illustrate the directional semantics of the case grid in use. (DATPOSSG) is a typical indirect object marked by the preposition 'to' – the man is the recipient, the goal. The transfer of the thing given stops at the goal. (DATPOSSL) is the local case of the dative possessive parameter. It is the case for the object transferred, in this instance, the £100.

In order to demonstrate the use of case in semantic analysis, we suggest that the correlative representation of the verb 'take' be compared.

```
[TAKE]—
  (ACTG)—[MAN: *x]
  (DATPOSSS)—[D: Tybald].
```

In this context, the man *x takes as the agent. He is the active recipient, but still the goal. While Tybald, still the source, would have been the original possessor. Comparing these two reciprocal predicates, 'give' and 'take' shows how the combined roles of each of the parties to the promise

can be represented. Furthermore, it shows how the transitivity of the verb's actions is clearly indicated. The semantic connection between the two verbs is apparent. The direction of the property transfer is changed but the representation allows the expression of the man's role as goal or receiver in both, and D's role as the source in both.

In figure 2, the judge's reasons for judgement in the same case show a little more of the expressive capability of the case-based representation.

In the proposition prefaced by the modal (JD) indicating the judge's decision, the judge is giving an opinion on a matter of fact. The modal indicates the context within which it is true, that is the possible world of his own opinion. His opinion is that there was no intention to contract. The negation symbol (–) in the label field indicates there was no intention to contract in that context. Similarly, a little further on, there was no instance of an event of being legally binding [LEGAL_BIND] in the context. The evidence (EVID) for there not being any intention to contract is the promise we saw above. The recipient of the promise was not known, [PERSON: ?]. The judge is saying that the statement of promise was too general. Since this fact conflicts with the assertion above, that the promise was made to the man *x. Again the [REASONS], 'warrants' in Toulmin's terminology, are within the scope of the modal (JD) expressing the judge's opinion.

A little further on there is another predicate of interest because of the case analysis.

The event [INTEND] involves a psychological

[REASONS: #1]→(INCL)→[(JD)–
[INTENTION_TO_CONTRACT: –]–
  (ACTS)→[D: Tybald]
  (EVID)→[PROMISE-α: #W1]–
  (CHRC)→[PHRASE: "GENERAL WORDS"],
[PROMISE-v: #W1]–
  (ACTS)→[D: Tybald]
  (DATPOSSG)→[PERSON: ?],
[LEGAL_BIND: –]–
  (ACTP)→[PROMISE-α: #W1]
  (OBJL)→[D: Tybald]
  (AMBS)→[INTEND: #W1]–
    (ACTS)→[D: Tybald]
    (DATPSYL)→[INTENTION_TO_CONTRACT: –]
        [INTENTION_TO_EXCITE: #W1]–
          (DATPSYG)→[SUITORS: {*}?]...]]

Figure 2. Reasons in *Weeks* v. *Tybald.*

[INTEND: #W1]–
  (ACTS)→[D: Tybald]
  (DATPSYL)→[INTENTION_TO_CONTRACT: –]
    [INTENTION_TO_EXCITE: #W1]–
      (DATPSYG)→[SUITORS: {*}?]..

verb. Nevertheless, the person intending is designated the agent as there is a strong element of will expressed in the verb. Especially in cases of this kind, where the motives of the parties are the subject of investigation, it is important to recognize and express the volitive element. Somers's cases make it possible to do it well. The content of the intent is the dative psychological local (DATPSYL) slot filler, which says that Tybald intended not to contract but to excite suitors. The suitors who will experience the excitement are designated dative psychological goal (DATPSYG) or experiencer. There is no volitive element in this event of excitation in contrast to the one of intention. The symbol in the label field of the suitors concept ({*}?) indicates that there is a set of unknown suitors.

Having seen how expressive the notation is for semantic analysis. We go on to show some more complex examples.

A second short case, *Stamper* v. *Temple* (1845 6 Humph. 113 (Tennessee)), has a similar fact situation but includes some psychological predicates and hypothetical concepts.

> TURLEY, J.: "'We are constrained to believe that what is called an offered reward of $200. was nothing but a strong expression of his feelings of anxiety for the arrest of those who had so severely injured him, and this greatly increased by the distracted state of his own mind, and that of his family; as we frequently hear persons exclaim, 'Oh, I would give a thousand dollars if such an event were to happen or vice versa'. No contract can be made out of such expressions: they are evidence of strong excitement, but not of a contracting intention."

Turley, J. hedges before stating his opinion. We do not represent the full expression of his doubt, for example, 'constrained', 'to believe', 'what is called...'. We do show that the promise is, in his opinion, not an offer, and so concentrate on the informational content of the report to the exclusion of the rhetorical style.

Here, the reward for the arrest of miscreants seems more likely to be a true offer, made in the spirit of vengeance than does the statement in the *Weeks* case. *Stamper* presents a very similar problem from a slightly different perspective. The

promise is represented in the same way as the one in the *Weeks* case and is not shown here. However, the additional psychological complication is of interest.

```
[STATE_OF_MIND: #S1]-
   (DATPSYG)→[D: Temple] [FAMILY: Temple]
   (AMBL)→[DISTRACT: #S1].
```

[STATE_OF_MIND] is a psychological predicate in which Temple and his family are conjoined as experiencers, dative psychological goal (DATPSYG), of the event. The ambient local (AMBL) case, a peripheral case describes a condition under which the predicated state prevails. The semantic content of the concept [DISTRACT] is the significant part. The syntactic relation conveys little more than would a conceptual relation such as 'attribute'. The ambient local case (AMBL) indicates a condition prevalent with regard to the atmosphere or locality of the predicate.

Further along in the representation, the (JD) modal operator makes a 'possible world' of the entire [REASONS] section of the case. The meaning of the passage rather than the linguistic content has been represented. We see . that in the opinion of Turley, J. the promise is not an offer. It was motivated by Temple's distraught state of mind, his anxiety.

```
[[PROMISE-n: #S1]-
   (-(EQUIV)→[OFFER: #S1]
     (CAUS)→[ANXIOUS_FOR: #S1]
        [STATE_OF_MIND: #S1].]
```

The conceptual relation cause (CAUS) is not a deep case. The only causation case the grid provides is the ambient source (AMBS) described as 'reason'. We did not judge D Temple's reason for making the promise to be his anxiety, but rather the injury. We, therefore, attempted to represent the relationship as a directly causative one, with its own specific conceptual relation, having semantic, but no syntactic content.

The promise is said not to be an [OFFER] but instead, has the characteristic (CHRC) of being an "expression of strong feeling". The descriptive phrase must be distinguished from the conceptual version. The two representations of the characteristic follow, the descriptive first, the conceptual next.

```
[[PHRASE:
   "EXPRESSION OF STRONG FEELING"]-
     (EQUIV)→[EXPRESSION: #S1]-
        (OBJG)→[FEELING: #S1]-
           (ATTR)→[STRONG: #S1]„„]
```

6. Retrieval

Information is retrieved by matching frames. Each frame is a list of symbolic expressions and they can be unpacked one after the other. Frames allow each item in the kb to function as an index key. There is no need to specify eligible keys ahead of time. In our proposed intelligent retriever, every concept is accessible.

6.1. LOG+

Miezitis's LOG (Lexical option generator) (1988) amplified as LOG+ is the algorithm used as the exemplar. When LOG is given a set of concepts, it can produce a selection of lexical options with similar meanings for each input concept. The matcher can cope with idiomatic phrases like legal concepts. Smart markers convey messages throughout the type hierarchy. In this way, complex expressions are matched to retrieve information. In the event of partial matches, the markers enable matched nodes to share information with other nodes. Each LOG node has a situation template containing restrictions on the hierarchical inheritance of properties. It also has a syntactic template that contains structure information and the lexical options available when a situation is matched.

Each search begins at the top of the hierarchy. Marker passing is constrained by a choice of most likely paths at an early stage. Generic nodes are marked 'most likely to succeed' or 'magnetized'. As soon as the process encounters nodes in which some required slots are matched, those nodes are marked, and the search becomes localized and intensified. Semantic constraints, the above mentioned selection rules, indicate which slots are required to be filled.

The description of LOG+ includes some other capabilities. In addition to negotiating matches within the is-a hierarchy, it also can handle partonomic matches. It can deal in a simple way with some contextual matches, and it has more facility with speech acts and metaphors. It can deal with matches involving the hypotheticals proposed by the judges. Most important, it is more flexible in matching and allows constrained access to subsumer options, in order to decide which lexical option is the appropriate match. LOG+ was so enhanced to allow its adaptation from lexical matching to concept selection.

6.2. A sample search

A very brief version of the description of a sample search follows. The question was derived from the fact situation in the case of *Carlill* v. *Carbolic*

Smoke Ball Co. (England. Court of Appeal. [1893] 1 Q.B. 256). A summary of the facts follows.

> D advertised a medicine ball in a local newspaper, offering £100 reward if cold or influenza followed upon its use in the prescribed manner. The offer was backed by a deposit of the stated amount in a local bank. P read the advertisement, bought a smoke ball, used it as directed, contracted influenza and sought to recover the £100.

The facts have been represented in the same way as those in the cases in the kb.

The first question asked regarding the *Carlill* case is whether or not the promise is a 'mere puff'.

> "Is this promise a mere puff?"
> [PROMISE-v: #Q1] = [MERE_PUFF]

An attempt is made to match the concept [MERE_PUFF], which fails. The system does not 'know' the concept, nor does it 'know' [PUFF]. A second attempt is made by providing a description of the concept. The description of the concept used as a question (figure 3)Figure is derived from a discussion of the mere puff in an introductory contract law text (Treitel 1979, p. 107, 244).[4]

In attempting to match the description we will find the promise in *Weeks* that was described by the judge as having been stated in 'general words'. [TERM] and [WORD] come together in the type hierarchy as near synonyms under the common parent, [LEXICAL_UNIT]. They will inherit the same principal attributes. There is nothing in the context of either concept to distinguish it

---

[PROMISE-n]–
(CHRC)→[TERM]→(ATTR)→[GENERAL] or
(CHRC)→[VAGUE] or
(CHRC)→[OPINION] or
(CHRC)→[PROMISE]→(ATTR)→[–SPECIFIC] or
(CHRC)→[PRECISE][–SERIOUS] or
(CHRC)→[ASSERTION: [•]]–
  (OBJG)→[FACT: [•]]→(ATTR)→[VERIFIABLE]
(–EVID)→[INTENTION_TO_CONTRACT]
(–EQUIV)→[OFFER],.

Figure 3. A mere puff.

---

[4] "A statement inducing a contract may be so vague, or so clearly one of opinion, that the law refuses to give it any contractual effect." (Treitel 1979, p. 107). "These are statements which are so vague that they have no effect at law or in equity.... The distinction is between indiscriminate praise, and specific promises or assertions of verifiable facts." (Treitel 1979, p. 244).

from the other. It is possible, therefore, to match [TERM] against [WORD]. Although the phrasal description of the promise in *Weeks* does not match exactly the first (CHRC) of [MERE_PUFF], a partial match will be reported as any type [PHRASE] will have its type label subjected to a character-by-character match. From the concept [WORD] in the hierarchy, there is a pointer to the use in the [PHRASE]. Furthermore, it is to be expected that [VAGUE] would be associated in a full lexicon with 'general' and 'meaningless'. The natural language usages have not been included in our type hierarchy here as they are not the subject of investigation, but it may be assumed that this term would be matched in the normal operation of a system in this domain. The phrases 'vague terms' and 'general words' would be matched as synonymous.

We find that a mere puff is *not evidence* of intention to contract and is not equivalent to an offer. In *Weeks*, the promise is evidence that there is *no intention to contract*.[5] The matcher can reason well enough to relate the variant negative structures. The promise in *Weeks* is a mere puff.

If we are to argue that the *Carlill* promise is not a puff, we will have to show how it is unlike them. One of the (CHRC)s of mere puff is that the promise is [–SPECIFIC]. There is no meaning defined for the concept, not specific. However, we do know that one of the judge's reasons in *Weeks* for deciding that the promise was not legally binding was that Tybald promised to [PERSON: ?]. Similarly, in *Carlill* the promise is made in an advertisement in a newspaper. The statement in *Carlill* appears to be as general as the one in *Weeks*. In our representation, it appeared in the consequent of the promise as follows: The person *x was the same unknown person who in the antecedent of the statement would buy a smoke ball and use it as directed in the advertisement. Note that the dative possessive path case (DATPOSSP) is used for the reward rather than the local case (DATPOSSL), which we saw used earlier for a possessory transfer of an object. The path case is

[PAY]–
  (ACTS)→[D: Carbolic]
  (DATPOSSG)→[PERSON: *x]
  (DATPOSSP)→[REWARD]–
    (MEAS)→[MONEY: @£100],
  (DATPOSSS)→[MONEY: @£1000]–
    (OBJG)←[DEPOSIT-v: #Q1],.

---

[5] *Weeks* is relevant as Treitel cites it as an example of mere puff (1979, p. 107).

used for a payment. It indicates a passage from the source to the goal usually. The source case here (DATPOSSS) tells us the reward money in question was the money deposited in the [DEPOSIT-v: #Q1] event, of which more later.

In the same way that *Weeks* was retrieved, so was the similar *Stamper* case found. The promise in *Stamper* is very similar to the promise in *Carlill*. Recall that the judge stated that Temple's promise was not equivalent to an offer because it was caused by his anxious state of mind. However, the promise in *Carlill* is *not* the product of emotion. It is instead a serious business proposition.

Something quite different about the *Carlill* promise is the designation of the source of the reward. The (DATPOSSS) slot includes a frame for the event [DEPOST-v: #Q1] as shown above. The representation of that concept makes the difference in the *Carlill* case. Our kb does not know what [DEPOSIT] is. When there is no match for a concept like [DEPOSIT], the first attempt to reconcile the conflict will involve generalizing to the next higher node. In this situation that is [GIVE]. [DEPOSIT] has the same syntactic characteristics as [GIVE] and does match the required slots for that concept. [DEPOSIT] might at some point be added as a useful specialization of [GIVE], or it might just be left as a partial match that is reported to the user. The decision depends on the need the implementation is intended to satisfy.

In the course of attempting to match the promise in *Carlill*, the LOG+would have trouble with [CONTRACT-v].

    [CONTRACT-v: #Q1]–
    (DATPOSSG)→[P: *Carlill*]
    (DATPOSSL)→[INFLUENZA: #Q1].

It has a (DATPOSSG), representing the benefactor, and a (DATPOSSL) slot, indicating the object that the benefactor receives, here, influenza. The semantic constraint in the knowledge base for the sense of [CONTRACT-v] about which we know specifies only an agentive role, (ACTS), which may be repeated. Within our kb, it is represented only as an intransitive verb (Vi). It is clear that the sense of 'contract' used in the question derived from *Carlill* is something different. Ideally the

    [DEPOSIT-v: #Q1]–
    (ACTS)→[D: Carbolic]
    (LOCL)→[BANK: Alliance]
    (ACTG)→[SHEW: #Q1]–
        (ACTS)→[D: Carbolic]
        (OBJL)→[SINCERE: #Q1].

new information would be acquired by the system. LOG+ has allowed for an inventory to be developed in association with failed searches. This new sense of contract, along with its slots, would become a part of that inventory list; the knowledge of 'contract' in the sense of 'contracting an illness' would be acquired by the system. If other instances of this use of the verb were added, they might be appropriately clustered. However, since it is unlikely that there will be such occurrences, this example might be saved as an exception that is potentially useful in establishing the meaning of unusual occurrences of [CONTRACT-v] in incoming queries.

Another problem arises with the attempt to match [SHEW], an archaic form of the verb 'to show' which is sometimes used in law reports. The alternative choice of [SHEW] for 'show' is another example of the use of near synonyms. Within the system, the words are considered to be exactly equivalent and are so represented as alternative lexical choices for the same conceptual representations, the same meanings. They will be adjacent nodes on the same level of the hierarchy, the ideal situation for synonyms. If terms suitable for use as synonyms are too widely separated, LOG+ will not find them because of the magnetization of the generic nodes early in the process to channel the search. If 'shew' for some linguistic reason had been placed far away from 'show', perhaps in some section of the hierarchy having to do with obsolete terms, it would have been missed in this search.

Money is deposited to 'shew sincerity', as a source (DATPOSSS) of reward money.[6] Furthermore, the fact that the money has actually been deposited in the Alliance Bank makes a verifiable fact of that part of the promise.

In the type hierarchy, 'serious', and 'sincere' are treated as if they were synonyms. Both are commonly applied to business matters within the same context. The offer here is sincere; there is a demonstration of intention to contract since the reward money has been deposited. The conceptual representation for [OFFER] in our lexicon did indeed fit the [PROMISE] in *Carlill*.

Finally, we see that a mere puff, even if it happens to be 'precise', is not to be 'serious'. And that it is not, whether vague or precise, evidence of an intention to contract and it is not an [OFFER]. It is clear that the *Weeks* and *Stamper* cases would

---

[6]'Reward' here is actually a compensation, but is indeed called 'reward' in the case report.

be retrieved in an attempt to determine what a mere puff is. Both cases contained propositions that were evidence that there was no intention to contract. Here we are looking for assertions that are not evidence of intention to contract. LOG+ is capable of reconciling the variant negatives in the course of the search. We have been able to match easily the [INTENTION_TO_CONTRACT: ~] cases.

The term 'mere puff' is a relatively modern legal concept. Its use dates from the time of *Carlill*. It is possible to find cases on 'new' legal concepts if their meaning is appropriately represented. At this point, the definition of the legal concept 'mere puff' could be added to the lexicon for future use.

7. Conclusion

Retrieval takes place on three interacting planes. The conceptual representation constitutes one plane, the lexicon another, and the argument structure a third. The arguments have been used to structure the discourse, however, the ultimate retrieval system would emulate legal reasoning. As the lawyer sought to develop his or her own new argument, he or she could review the arguments of others on similar issues and use or discard them as he or she wished. This representation is the underlying preparation for that further development.

It is assumed that in due course, systems of this type will be able to communicate with their users in natural language at both the question and answer ends; that is, both language analysis and language generation by machine will be possible. At present, the representation used describes the meaning of the text adequately for inference at a suitable level. It is not adequate to make possible a full translation from English, nor the generation of a full text response in English. We are attempting to model conceptual content in order to facilitate the retrieval of information rather than to reason to definitive conclusions. It has been demonstrated that retrieval based on semantics and inference can be perceptive and powerful.

References

Dick, Judith P (1991). *A conceptual case-relation representation of text for intelligent retrieval.* PhD dissertation, forthcoming, August 1991. Toronto: University of Toronto, 1991.

Fillmore, Charles J (1968). "The case for case." In Bach, Emmon Werner and Harms, Robert Thomas (editors) (1968). *Universals in linguistic theory.* New York: Holt, Rinehart and Winston, 1968. 0–88 [sic].

Hirst, Graeme (1987). *Semantic interpretation and the resolution of ambiguity.* Cambridge: Cambridge University Press, 1987.

Holmes, Oliver Wendell (1897). "The path of the law." In MacGuigan, Mark Rudolph (editor) (1966). *Jurisprudence: Readings and cases.* Toronto: University of Toronto, 1966, 48–62.

Karlgren, Hans (1977). "Challenge paper: Homeosemy—on the linguistics of information retrieval" In Walker, Donald E and Karlgren, Hans (editors) (1977). *Natural language in information science: Perspectives and directions.* (International Federation for Documentation FID Pub. 551) Stockholm: Skriptor, 1977.

Meadow, Charles T (1979). "Information science and scientists in 2001." *Journal of information science,* 1 217–222.

Miezitis, Mara Anita (1988). *Generating lexical options by matching in a knowledge base.* Technical Report CSRI-217. Computer Systems Research Institute, University of Toronto, September, 1988.

Milner, James Bryce (1985). *Milner's cases and materials on contracts.* 4th ed. Edited by S M Waddams. Toronto: Emond Montgomery, 1985.

O'Connor, John (1980). "Answer-passage retrieval by text searching." *Journal of the American Society for Information Science.* 31, 227–239.

Schank, Roger Carl; Kolodner, Janet L and De Jong, Gerald (1981). "Conceptual information retrieval." In Oddy, Robert Norman, *et al.* (editors) (1981). *Information retrieval research.* London: Butterworths, 1981, 94–116.

Somers, Harold L (1987). *Valency and case in computational linguistics.* Edinburgh: Edinburgh University Press, 1987.

Sowa, John F (1984). *Conceptual structures: Information processing in mind and machine.* Reading, MA.: Addison-Wesley, 1984.

Toulmin, Stephen Edelston (1958). *The uses of argument.* Cambridge: Cambridge University Press, 1958.

Treitel G H (1979). *The law of contract. 5th ed.* London: Stevens & Sons, 1979.

	Source	Path	Goal	Local
Active	instigator of action +/-volitive +/-animate	instrument or means	intended result (-animate) active recipient (+animate)	non-passive patient
Objective	original state (-concrete) material (+concrete)	counter-instrument passive means	result state (-concrete) factitive (+concrete)	undergoing change-of-state
Dative psychological: possessive:	stimulus original owner	medium medium/price	experiencer +/-dynamic recipient	content thing transferred
Locative	place from where	space traversed	final destination	static position
Temporal	time since	duration	time until	time at which
Ambient	reason	manner	aim(+volitive) consequence (-volitive)	condition

Table 1. Somers's case grid (1987, p. 206).