Representation of legal text for conceptual retrieval

Judith P. Dick

1. Introduction

State-of-the-art information retrieval provides documents by means of Boolean search using keywords (Salton and McGill 1983). Conceptual information retrieval provides information about a given concept or concepts from a knowledge base by means of frame matching. The research reported here is an attempt to move from state-of-the-art information retrieval to conceptual retrieval.

Oliver Wendell Holmes demonstrated the need for conceptual information retrieval in law with 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 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.)

The Vermont justice of the peace was looking for information on a concept behind the facts he could not name. As Holmes commentary indicates, our commonly used tools of access to case law would force one to look for concepts under catch-word type subject headings. And, as he points out, such headings sometimes result in burying concepts in unfortunate categories or under arbitrary names. With a conceptual retrieval system, one does not have to look for law under a term such as 'churn' or conjure up the name of the appropriate subject heading. One is able to retrieve information on the concept or concepts involved, if he or she provides a variant of its usual name, or some description, or even a situational context. The user is able to locate any information about the concept(s) in his query that is stored in the knowledge base at hand. Conceptual retrieval is exact: it homes in on exactly the information sought. There is no need to choose between emphasis on precision or on recall. Anything that answers the question must be recalled. Nothing that is extraneous may be included.

Carole Hafner described a conceptual retrieval system in 1978. Other AI research has focussed on legal reasoning and although it has a bearing on conceptual retrieval (in particular the work of Gardner (1987) and of Rissland and Ashley (1987) and Ashley (1991)) it has not attempted to come to terms with text analysis, a major stumbling block in developing conceptual retrieval systems. This paper describes work directed toward cutting that block down to size.

2. Overview

We contend that the retrieval of conceptual information from legal text is dependent upon the construction of a viable knowledge representation. Furthermore, we contend that the concepts to be represented must be derived from the text if retrieval is to be meaningful. If a selection of pre-assigned concepts, that is, subject headings, were to be related to the text, we contend that the resulting system would share a number of difficulties with state-of-the-art retrieval systems, even if the pre-assigned concepts were fully represented. Such an approach would produce a deluxe index, but would be unlikely to produce the conceptual retrieval we seek.

In our present state of knowledge, a suitable knowledge representation to be used for information retrieval, must be coarse-grained. In this research, a coarse-grained representation has been developed using caseframes that describe objects derived from the text itself. The cases are taken from Harold Somers's grid of cases especially developed and recommended for use in the machine analysis of language (Somers 1987). Somers's view of case is based on an original interpretation of valency.

In order to attempt a realistic approach, the work is modelled on the activity of lawyers and legal researchers in retrieving the information they need. The goal of their information seeking is to build an argument to answer the problem at hand. In developing the knowledge representation, each case was analyzed so as to make its argument readily available to the user. It is a natural approach and leads directly to a method for searching for argument-oriented information. Issues are analyzed in terms of the interaction of facts and legal concepts. Within the representation of each argument, the facts are separated from the reasons for judgement, making it possible to answer questions about different kinds of issues without ambiguity.

There is a lexicon of legal concepts. Each has its own definition. The definitions have been taken from legal dictionaries, treatises and other authoritiative sources. As well, each legal concept has 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 are allowed to develop naturally as knowledge accumulates. They do not have single dimension, rigid definitions, but may present meaning from numerous perspectives.

The objective of retrieval is to access information about what a concept, or combination of concepts, actually entails. We can expect to retrieve information that is implicit in the knowledge base. Within the system, the meanings of relations are fully specified as a part of the conceptual analysis. The representations allow us to disambiguate meanings that might otherwise be troublesome. It is possible to distinguish the concepts entailed by frequently used terms such as "court" and "trial" that are difficult to use in keyword retrieval. Partial descriptions of the concepts or the situations that involve them typically can be used in successful retrieval. It is possible to successfully retrieve information related to some abstract concepts. The degree of success depends upon the clarity of the information and the quality of the representation.

3. The elements of the representation

The knowledge base consists of the representations of contract cases. The cases were taken from a casebook (Milner 1985). It was important that the group of cases share enough common concepts to make demonstrations of some reasonable matches and conceptual comparisons.

The principal argument of each case was analyzed according to Toulmin's argument model so that the elements of each argument were separated and analyzed individually (Toulmin 1958). The Toulmin or "good reasons" model of analysis is intended to be applied to any argument type in any domain. It is an uncomplicated model and is used here as a schema structuring the discourse. Toulmin starts with the 'claim' of the argument, the goal, as in fact one often does in making an argument. He then proceeds to relate the remaining parts of the argument to the 'claim'. The 'data' or 'grounds' are the facts mashalled in support of the 'claim'. The 'warrant' is the link between the 'grounds' and the 'claim'. It is a hypothetical bridge-like statement (Toulmin 1958, p. 100). In the representations, reasons for judgement are placed in this category. Next comes the 'backing' or authority for the argument and the 'modal qualifiers' which limit its scope. Finally, any countervailing argument is classed as a 'rebuttal'.

Sowa's conceptual graphs are the notation used for the representations and they provide the logical base of the system (Sowa 1984). A conceptual graph is a finite, connected, bipartite graph (Sowa 1984, p. 73). Formally, a graph G consists of a nonempty set N, of *nodes*, and a set A, of *arcs* between the nodes (Sowa 1984, p. 375). A conceptual graph and the Peano-Russell notation (a frequently-used notation for first-order logic) for the same assertion are shown in Figure 1.

The cat sat on the mat.
An English sentence

[CAT]→(STAT)→[SIT]→(LOC)→[MAT]
A simple conceptual graph

∃x∃y∃z(cat(x) and stat(x,y) and sit(y) and loc(y,z) and mat(z))
Peano-Russell notation

Figure 1. Examples of notation.

The square brackets indicate concept nodes, and parentheses indicate conceptual relation nodes. The arrows that link the conceptual relations to the concepts represent arcs. The linear, rather than the original graphic, form of the notation is used here as it accommodates variables, is closer to commonly-used first order logic notation and is well-adapted for use in sentential analysis. The notation is expressive. For example, the method of limiting the scope of quantifiers can be seen in Figure 2.

[MAN: ∀]←(EXPR)←[LOVE]→(PTNT)→[WOMAN: ∀]
Every man loves every woman.

[MAN: ∀]←(EXPR)←[LOVE]→(PTNT)→[WOMAN]
Every man loves some woman.
(For each man, there exists some woman that he loves.)

[WOMAN: *x][[MAN: ∀]←(EXPR)←[LOVE]→(PTNT)→[*x]]
Every man loves some woman.
(There exists a woman x whom every man loves.)

[MAN: ~*x] [[*x]←(EXPR)←[LOVE]→(PTNT)→[WOMAN: ∀]]
There is no man who loves every woman.

Figure 2. Quantifier scoping, Sowa 1987, p. 13-14.

Conceptual graphs are mathematically well-founded, but their greatest advantage here, is that they are especially useful for work in natural language understanding, as Sowa stresses the importance of semantic interpretation (Sowa 1984, p. 211). Furthermore, he has not tied the graphs to a particular linguistic theory, but instead has demonstrated their versatility. Of particular interest in the context of this research, he has recognized the traditional *deep cases* including 'agent', 'patient', 'instrument' and so on, as conceptual relations. Case theory was introduced by Fillmore in 1968. 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)

Somers's grid of cases has been used to supplement Sowa's conceptual relations. Somers specified a set of twenty-eight cases with clearly defined relations that combine semantic realizations with grammatical relations. In

245

so doing, he was able to answer the most ringing criticisms of case theory, including the problem of source-goal directionality. He provided solutions for a number of dual-role problems in case assignment that had resulted in a haphazard proliferation of cases. It was apparent from the first that they were particularly suitable for use in this research.

4. Examples of the knowledge representation

The knowledge base consists of a lexicon of legal concepts and the representations of arguments from the cases. Excerpts from two simple cases are shown below in order to display aspects of the representation. In the case of *Weeks* v. *Tybald*,[1] the text of the case is quite simple,

> "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 representation using conceptual graphs and Somers's cases appears in Figure 3. This representation shows an argument and demonstrates the use of the argument schema and the caseframes employing Sowa's notation and Somers's cases. The [CLAIM], the first section of the [ARGUMENT], states that no contract actually existed, and that the defendant (D) is not legally bound. [~CONTRACT-n: #W1] designates an inchoate entity, something that is without existence, but something which we discuss, as we might a unicorn. It may be an entity of some kind, but it is not a contract. The negative [LEGAL_BIND: ~] in contrast indicates that there is no occurrence of legal binding in this context. Further, [INTENTION_TO_CONTRACT: ~] means that there was an absence of intention to contract rather than that there was an intention *not* to contract. Evidence of the fact that D had not intended to contract is the promise he made. Determining whether or not D intended to do something is a central issue. Intention as it is used in the concept of 'intention to contract', means a definite, willful decision. If intention were to be treated as a modal operator, it would create a possible world in which the intention existed, taking the main factual issue of the case out of the mainstream of the argument. If 'intend' were to be treated as a psychological verb, D would not be the subject but the experiencer, in Somers's terms, the dative psychological goal (DATPSYG) case. The important element of volitive decision making would not be expressed. For these reasons, the concepts deriving from the verb 'to intend' have case assignments like those associated with a verb of action. (ACTS) the active source case, indicates agency and volition on the part of the subject.

---

[1] (1605) Noy 11; 74 E.R. 982.

The [GROUNDS] follow directly after the [CLAIM]. Conjunction is represented by continguous graphs, not by a specific symbol. [GROUNDS] contain the factual content of the case. A promise is described. [PROMISE-v: #W1] is the instance and D Tybald is the instigator. The recipient is either the plaintiff (P) Weeks or his father. It appears, according to the text, the promise was made to Weeks or to his father. Putting aside the question of truth values, the problem of representing a disjunction is complex. The text is not precise, so the representation cannot be precise. The promise may have been made to the son, the father, to either, or to both. Furthermore, the statement is preceded by "it would appear", adding to the uncertainty. The scope of the disjunction is indicated by the slot (DATPOSSG).

The text does not say 'promise' but 'told'—the judge is simply relating the facts of the situation. In the representation some structure has been added for promises in order to present similar constructions with regularity. They are all in the form of conditional statements.

Within the context of the promise are a number of nested graphs, contexts within contexts. The conceptual relation contains, (CONT), the concept [TERMS]. [TERMS] is a subtype of the concept [SITUATION] that groups together propositions with similarities. [TERMS] is used to keep together a number of graphs within the context of the promise.

The antecedent of the implication is a conjunctive concept, 'to marry with consent'. D Tybald says that if some man, *x, marries his, Tybald's, daughter and Tybald himself consents to the marriage, then Tybald will give the man in question the amount of £100.

[CONSENT_TO] is a particle construction in which the preposition is kept with the verb in order to achieve an accurate semantic representation of the verb phrase, as if it were an idiomatic phrase. This is one of many examples of verb phrases that cannot be decomposed. This solution is not entirely satisfactory.

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. The implicitly conjoined phrase 'aver and declare' is typical of legal syntax. Such a phrase commonly becomes an indexing term. The two verbs are used together for emphasis. It functions as an indication that a legal principle of serious import is at hand. Here, the phrase may indeed indicate that the judge is citing the pleadings as problematic. The judge is saying that neither by action nor by words did D indicate intention to contract.

The judge goes on to say that D did not indicate someone as the recipient of the promise. Since this fact conflicts with the assertion above, it is significant that it is distinguished as the judge's belief.

```
[ARGUMENT: #1]→(INCL)→[
[CLAIM: #1]→(INCL)→[
    [~CONTRACT-n: #W1]–
        (PARTY)→[P: Weeks]
        (PARTY)→[D: Tybald].
    [LEGAL_BIND: ~]–
        (ACTP)→[PROMISE-n: #W1]
        (OBJL)→[D: Tybald]
        (AMBS)→[INTENTION_TO_CONTRACT: ~].
    [INTENTION_TO_CONTRACT: ~]–
        (ACTS)→[D: Tybald]
        (EVID)→[PROMISE-n: #W1].  ] ;end of CLAIM
[GROUNDS: #1]→(INCL)→[
    [P: Weeks]→(EQUIV)→[MAN: Weeks]
    [PROMISE-v: #W1]–
        (ACTS)→[D: Tybald]
        (DATPOSSG)→[P: Weeks] or [FATHER: Weeks]
        (DATPOSSL)→[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: @L100],]
        (TEMPL)→[TIME: promise#W1].
    [MARRY: #W1]–
        (ACTS)→[P: Weeks]
        (ACTL)→[DAUGHTER: Tybald]
        (TEMPL)→[TIME: marry#W1-]→(>)→[TIME: promise#W1+].
    [CONSENT_TO-v: #W1]–
        (ACTS)→[D: Tybald]
        (OBJL)→[MARRY: #W1].
    [~[GIVE: #W1]–
        (ACTS)→[D: Tybald]
        (DATPOSSG)→[P: Weeks]
        (DATPOSSL)→[MONEY: @L100].]
        (JD)→[[~AVER][~DECLARE]–
                (ACTS)→[D: Tybald]
                (OBJG)→[PERSON: name]–
                        (DATPOSSG)←[PROMISE-v: #W1],.] ] ;end of GROUNDS
[REASONS: #1]→(INCL)→[
    (JD)→[[INTENTION_TO_CONTRACT: ~]–
                (ACTS)→[D: Tybald]
                (EVID)→[PROMISE-n: #W1]→(CHRC)→[PHRASE: "GENERAL WORDS"].
            [PROMISE-v: #W1]–
                (ACTS)→[D: Tybald]
                (DATPOSSG)→[PERSON: ?].
            [LEGAL_BIND: ~]–
                (ACTP)→[PROMISE-n: #W1]
                (OBJL)→[D: Tybald]
                (AMBS)→[INTEND: #W1]–
                        (ACTS)→[D: Tybald]
                        (DATPSYL)→[INTENTION_TO_CONTRACT: ~]
                            [INTENTION_TO_EXCITE: #W1]–
                            (DATPSYG)→[SUITORS: {*}?],,.]]] ;end of REASONS, end of ARGUMENT
```

Figure 3. *Weeks* v. *Tybald*, (1605) Noy 11; 74 E.R. 982.

It also highlights the problem referred to concerning the representation of disjunctions. The judge is stating that the statement was too general. Again the [REASONS], 'warrants' in Toulmin's terminology, fall within the modal (JD)

expressing the judge's opinion. In fact, there is no real argument here, but a decision on the facts. The schematic analysis makes the relationships of the propositions stand out clearly.

```
[ARGUMENT: #2]→(INCL)→[...
[GROUNDS: #2]→(INCL)→[...
   [STATE_OF_MIND: #S1]-
      (DATPSYG)→[D: Temple] [FAMILY: Temple]
      (AMBL)→[DISTRACT: #S1].
   [PROMISE-v: #S1]-
      (ACTS)→[D: Temple]
      (DATPOSSG)→[PERSON: {*}?]
      (DATPOSSL)→[PROMISE-n: #S1]→(CONT)→[TERM:
                    if [ARREST-v: #S1]-
                       (ACTP)→[PERSON: *y]
                       (OBJG)→[PERSON: Col{*}*x],
                    then [GIVE: #S1]-
                          (ACTS)→[D: Temple]
                          (DATPOSSG)→[PERSON: *y]
                          (DATPOSSL)→[[REWARD: #S1]→(MEAS)→[MONEY: @$200]],]....] ;end of GROUNDS
[REASONS: #2]→(INCL)→[
      (JD)→[[PROMISE-n: #S1]-
              (~EQUIV)→[OFFER: #S1]
              (CHRC)→[[PHRASE: "EXPRESSION OF STRONG FEELING"]-
                        (EQUIV)→[EXPRESSION: #S1]-
                                  (OBJG)→[[[FEELING: #S1]→(ATTR)→[STRONG: #S1]] or [ANXIETY: #S1]]]
              (CAUS)→[ANXIOUS_FOR: #S1] [STATE_OF_MIND: #S1]
      [HYPO: [PROMISE-n: #S2]→(CONT)→[TERM:
             if [HAPPEN: #S1]-
                (ACTS)→[EVENT: *a]
             then [GIVE: #S2]-
                   (ACTS)→[PROMISOR: *m]
                   (DATPOSSL)→[REWARD: #S2]→(MEAS)→[MONEY: @$1,000]].
      [PROMISE-n: #S3]→(CONT)→[TERM:
             if ~[[HAPPEN: #S2]-
                (ACTS)→[EVENT: *b]]
             then [GIVE: #S3]-
                   (ACTS)→[PROMISOR: *n]
                   (DATPOSSL)→[REWARD: #S2]-
                                (MEAS)→[MONEY: @$1,000].]]]
      [PROMISE-n: #S1]→(COMP)→[PROMISE-n: #S2]
      [PROMISE-n: #S1]→(COMP)→[PROMISE-n: #S3]
      [PROMISE-n: #S2][PROMISE-n: #S3]→(EVID)→[EXCITEMENT: *]-
                                                (ATTR)→[STRONG: #S2]
      [PROMISE-n: #S2]→(~EVID)→[INTENTION_TO_CONTRACT: *]
      [PROMISE-n: #S3]→(~EVID)→[INTENTION_TO_CONTRACT: *]
      [PROMISE-n: #S1]→(EVID)→[INTENTION_TO_CONTRACT: ~]].. ;end of REASONS, end of ARGUMENT
```

Figure 4. Excerpts from *Stamper* v. *Temple*, (1845) 6 Humph. 113 (Tennessee).

The second small case, *Stamper* v. *Temple*[2] has a similar fact situation. Its analysis shows the operation of the schema; and it includes a number of psychological predicates and hypothetical concepts that are a bit different.

> 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."

Here, the reward for the arrest of miscreants seems more likely to be a true offer, made in the spirit of vengeance, than a casual statement by a father hoping to have his daughter married. The facts make *Stamper* appear to present the same problem from a slightly different perspective. Excerpts from the representation for *Stamper*, are shown in Figure 4.

[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 predicates 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'.

---

[2] (1845) 6 Humph. 113 (Tennessee)

248

The promise has the same structure as the one in *Weeks*, showing the consistency of the format. It is made to an unknown person or persons. In the antecedent, if some person, *y, would effect the arrest of the perpetrators of the injury he would be acting as an instrument, active path (ACTP), to bring about the arrest.

The consequent of the promise is dominated by the predicate [GIVE], as it was in *Weeks*, although the meaning is stronger. A reward is being offered. In a sense, 'give' is a euphemism for 'pay'. In the previous case, it is unlikely that Weeks would have substituted 'pay' in speaking to someone who would marry his daughter.

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. 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 judge states in the second line of the text that the statement had been characterized as an 'offered reward'. Later, he stated that it was nothing but an 'expression of strong feeling'. Although the judge did not say directly that the promise was not an offer, the underlying idea is apparent in his reasons.

The judge proposes two hypothetical [HYPO] promises. They may exist in the same world as the rest of the context, the rest of the (JD) propositions, or they may exist in another possible world. Both promises are included in the one situation ([HYPO] is a subtype of [SITUATION]) as they are closely associated. The judge proposes a comparison between D's promise and his hypothetical one. He is saying that D's promise is analogous to his hypothetical one. A relation compare (COMP) had to be defined for this use. Sowa uses a relation (COMP) in defining the relation (ABOV) (1984, p. 226), but does not include a definition of it in his work.

4.1. Complex cases

So far we have seen the representation of only simple cases. The method used lends itself to the representation of complex arguments as well. The case of *Upton-on-Severn Rural District Council* v. *Powell*[3] deals with the problem of a farmer, seeking the service of a fire brigade while his barn was burning. Unfortunately, the wrong fire brigade was called and he was compelled to pay for the service he received. The request described in Figure 5 shows the county court judge's description of the convoluted fact situation involving the unclear request. Assertions within the scope of the (CCJ) modal are attributed to the County Court Judge. They are not facts in the sense that they are true in the real world, but rather in the sense that they are

said to be true by the judge. Now we encounter the second description of the request. At the time of the phone call, Powell is now said to have requested that the police inspector send *the* fire brigade. (It was previously reported that he requested *a* fire brigade). Here, as in the actual conversation, the referent is not clear. We don't know which fire brigade is *the* brigade [FIRE_BRIGADE: #]. This use of the '#' symbol has been used here to stress the expression of *the* in the text. We know only that the judge has declined to use the generic term [FIRE_BRIGADE: *] (some or any fire brigade) that was used in the first instance. The police inspector summoned the 'local' fire brigade in a manner that was 'natural'. One of the most difficult matters in this representation is the conceptual analysis of 'local'. The judge describes the act of calling the local fire brigade 'natural' in the sense of doing the usual thing. Note that the (AMBP) case is used, which typically accommodates manner modifiers ending in '-ly'. The (CCJ) says that the inspector "took the order as being one for the fire brigade with which he was connected." In the representation, 'understand' is substituted for 'took as being' since there appeared to be no reason to preserve that expression as an ungainly predicate. [UNDERSTAND] is a psychological verb without a volitive element. In normal circumstances, the inspector's will would not have affected his ability to understand. This is the sense in which the conceptual representation of the verb is made here. The inspector is the (DATPSYG), the experiencer, of the event of understanding. He understands the (DATPSYL), the content of the idea, namely that *the* fire brigade is that one with which he is himself connected, that is to say, the Upton fire brigade.

A lambda expression is used to present [REQUEST-n: #U2], the request for *the* fire brigade, as the police inspector understands it, that is, from the CCJ's point of view. The expression functions as a transcription of a restrictive clause. However, the restriction to the [REQUEST]'s description is valid only within the context of the predicate [UNDERSTAND: #U1]. To some extent, this definition of his understanding of *the* fire brigade is also a conceptual representation of the idea of 'local'. It interprets 'local' to mean 'connected with the police inspector'. Finally, the bracket in the left margin signals the closing scope of the modal (CCJ).

---

[3] England. Court of Appeal. [1942] All E.R. 220.

249

(CCJ)→[[REQUEST-v: #U2]–
            (ACTS)→[A: Powell]
            (DATPOSSG)→[POLICE_INSPECTOR: #U1]
            (DATPOSSL)→[[REQUEST-n: #U2]→(CONT)→[[SEND: #U2]–
                                    (DATPOSSL)→[FIRE_BRIGADE: #],]]
            (TEMPL)→[TIME: telephone-v#U1].
      [SUMMON: #U1]–
            (ACTS)→[POLICE_INSPECTOR: #U1]
            (OBJG)→[[FIRE_BRIGADE: Upton]→(ATTR)→[LOCAL: #U1]]
            (AMBP)→[NATURAL: #U1].
      [UNDERSTAND: #U1]–
            (DATPSYG)→[POLICE_INSPECTOR: #U1]
            (DATPSYL)→[[REQUEST-n: #U2]=
                  lambda x [REQUEST-n: *x]–
                        (OBJG)→[[[FIRE_BRIGADE: #]–
                                    (ATTR)→[[CONNECT: #U1]–
                                                (OBJG)→[POLICE_INSPECTOR: #U1],],]

Figure 5. An excerpt from *Upton v. Powell*

[PROMISE-v: #Q1]–
      (ACTS)→[D: Carbolic]→(CONT)→[TERM:
                  if [USE]–
                        (ACTS)→[PERSON: *x]
                        (OBJL)→[CS_BALL]
                        (AMBP)→[AS_DIRECTED: #Q1]–
                                    (LOCL)→[CS_BALL: #],
                     [CONTRACT-v]–
                        (DATPOSSG)→[PERSON: *x]
                        (DATPOSSL)→[INFLUENZA],
                  then [PAY]–
                        (ACTS)→[D: Carbolic]
                        (DATPOSSG)→[PERSON: #]
                        (DATPOSSP)→[REWARD]–
                                    (MEAS)→[MONEY: @L100],,
                        (DATPOSSS)→[MONEY: @L1000]–
                                    ←(OBJG)←[DEPOSIT-v: #Q1]–
                                                (ACTS)→[D: Carbolic]
                                                (LOCL)→[BANK: Alliance],
                                                (ACTG)→[SHEW: #Q1]–
                                                            (ACTS)→[D: Carbolic]
                                    (OBJL)→[SINCERE: #Q1].
[DEPOSIT-v: #Q1]–
      (ACTS)→[D: Carbolic]
      (OBJG)→[MONEY: @L1000]
      (LOCL)→[BANK: Alliance].
[USE: #Q1]–
      (ACTS)→[P: *Carlill*]
      (OBJG)→[CS_BALL: #Q1]
      (AMBP)→[AS_DIRECTED: #Q1].
[CONTRACT-v: #Q1]–
      (DATPOSSG)→[P: *Carlill*]
      (DATPOSSL)→[INFLUENZA: #Q1].
[~[PAY: #Q1]–
      (ACTS)→[D: Carbolic]
      (DATPOSSG)→[P: *Carlill*]
      (DATPOSSP)→[REWARD: #Q1]–
                  (MEAS)→[MONEY: @L100],.]

Figure 6. *Carlill v. Carbolic Smoke Ball Co.*

250

5. Retrieval

Retrieval is accomplished by matching frames and parts of frames. Meizitis's LOG (1988) is the algorithm, which, with some additionsi, is used as the model. The matcher can cope with idiomatic phrases with internal variables and will report partial matches. Smart markers are used to convey message throughout the net in the event of partial matches, in order to share information with other partially matched nodes and so facilitate complex and rapid retrieval.

Each search begins at the top of the hierarchy. Marker passing is constrained by a choice of most likely paths at an early stage, by marking generic nodes, in order to restrict search and avoid silly mismatches. 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 contraints developed from the representation indicate which slots are required. From there, the direction of the search depends on the content of the question.

5.1. Demonstrating retrieval

In the case of *Carlill* v. *Carbolic Smoke Ball Co.*,[4] D advertised a medicine ball in a local newspaper, offering £100 reward if cold or influenza followed upon its appropriate use as prescribed. The offer was backed by a deposit of the stated amount in a local bank. P read the advertisement, bought a ball, used it as directed, contracted influenza and sought to recover the £100. A representation of the facts appears in Figure 6. Assuming that the problem of 'mere puff' was being searched before the case went to court, the question is whether or not the promise is a 'mere puff'.

"Is this promise a mere puff?"

[PROMISE-v: #Q1] = [MERE_PUFF]

The representation of [MERE_PUFF][5] appears in Figure 7.

In attempting to match the conceptual definition of [MERE_PUFF] we will find the promise in *Weeks* that was said to be stated in 'general words'. Both [TERM] and [WORD] come together as 'synonyms' under the common parent [LEXICAL_UNIT]. They will inherit the same principal attributes. There is nothing in either the context of the [MERE_PUFF] or the context of the promise in *Weeks* to distinguish the use of either. It is possible, therefore, to match [TERM] against [WORD]. Although the phrasal description of the promise in *Weeks* does not match

---

[4] England. Court of Appeal. [1893] 1 Q.B. 256.

[5] The definition is derived from a discussion of the concept in an introductory contracts text (Treitel 1979), where it is said that, "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). And later, "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).

---

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].

[MERE_PUFF]–
(DEFN)→[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 7. The concept of 'mere puff'.

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 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 taken as synonymous.

Furthermore, 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*.[6] The matcher can relate the two variant negative structures to produce matching meanings insofar as the promises are concerned.

The user wants to know if his [PROMISE] is a [MERE_PUFF]. We can compare his promise to those retrieved as mere puffs. If he is to argue that the *Carlill* promise is not a puff, he 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. 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. Once again, there is no (DATPOSSG), no individual recipient to whom the statement inducing a contract is made. There is little to make it clear in *Weeks* what else characterizes the [PROMISE] as 'general words'. So far, the statement in *Carlill* appears to be as general as the one in *Weeks*.

In the same way that *Weeks* is retrieved, so is *Stamper* found. The promise in *Stamper* is very similar to

---

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

251

Retrieval takes place on three interacting levels. The coarse-grained representation of text 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 own new argument, he could review the arguments of others on similar issues and use or discard them as he 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

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*].

GARDNER, Anne von der Lieth (1987). *An artificial intelligence approach to legal reasoning*. Cambridge, MA: MIT Press, 1987.

HAFNER, Carole D (1978). *An information retrieval system based on a computer model of legal knowledge*. Ph.D. Dissertation, University of Michigan, Ann Arbor, MI: University of Michigan, 1978.

HAFNER, Carole D (1987). "Conceptual organization of case law knowledge bases." *Proceedings, First International Conference on Artificial Intelligence and Law*. Boston, May 1987. New York: Association for Computing Machinery. 35–42.

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 Press, 1966. 48–62.

JACKENDOFF, Ray S (1972). *Semantic interpretation in generative grammar*. Cambridge MA: MIT Press, 1972.

MEIZITIS, Mara Anita (1988). *Generating lexical options by matching in a knowledge base*. [1] MSc Thesis, University of Toronto, September 1988. [2] 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.

PERELMAN, Chaim and OLBRECHTS-TYTECA, Lucie (1969). *The new rhetoric: A treatise on argumentation*. London: Routledge and Kegan Paul, 1969.

SALTON, Gerard and McGILL, Michael J (1983). *Introduction to modern information retrieval*. New York: McGraw-Hill, 1983.

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.

the promise in *Carlill*. There is even a specific reward offered. However, in the [REASONS], the judge states that the promise is not equivalent to an offer because it was caused by D's anxious state of mind. However, the promise in *Carlill* is *not* the product of emotion. It is instead a well thought out business offer.

Now, it might have been argued that, in stating the precise amount of the reward, D had avoided the possible claim that the promise was imprecise or not specific. However, the reward in *Stamper* is similarly spelled out, and that promise was not judged as a binding offer. In *Stamper* as well, no recipient is designated. The nature of the reward as a sort of published offer is repugnant to the judge as his hypotheticals indicate. These two [HYPO] promises would also be retrieved in the process of the match. Their true value is of course limited to the possible world delineated by the modal operators in whose scope they are found.

Something quite different about the *Carlill* promise is the designation of the source of the reward. This frame in the (DATPOSSS) slot cannot be matched within the knowledge base. However, there are several things to be noted about the representation as they would involve some difficulties in themselves in the matching process. In the course of matching the [PROMISE-v: #Q1], the matcher has some trouble with [CONTRACT-v]. It has (DATPOSSG), representing the traditional benefactor, here, a person, 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. It is, within our knowledge base, an intransitive verb. It is possible that the verb 'to contract' could be represented as having a 'factitive' type of object (OBJG), a [CONTRACT-n]; however, that use is not required and has not been made in this knowledge base. It is clear that the sense of 'contract' used in the question derived from *Carlill* is something different. Ideally the new information would be acquired by the system.

The knowledge base 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 will be [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.

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 'synonyms'. In this case, the words *are* considered to be exactly equivalent, and are so represented as alternative

lexical choices for the same conceptual representation. Note that they will be adjacent nodes on the same level of the hierarchy, the ideal situation for a synonym. If terms suitable for use as synonyms are too widely separated, LOG will not find them because of the magnetization procedure which primes a generic node early in the process to channel the search narrowly.

Money is deposited to 'shew sincerity', as a source (DATPOSSS) of reward money.[7] 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, it will be noted that 'serious', and 'sincere' are synonyms. They are commonly applied to business matters within the same sort of context of gravity and sobriety of action, that is, behavior that is 'serious' in the sense of 'not frivolous'. The point is that the offer here is sincere; there is a demonstration of intention to contract, since the reward money has been deposited. The concept, [OFFER-n], fits the [PROMISE] in *Carlill*.

Finally, we see that a [MERE_PUFF], even if it happens to be 'precise', is said 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]. We have been able to match easily the [INTENTION_TO_CONTRACT] cases. It is clear that they would be retrieved by this query as well.

Something new has happened in this case. The term 'mere puff' is a relatively modern legal concept. Its use dates from the time of *Carlill*. It would not have been derived from the early cases we have discussed so far, but, in standard indexing practice it might be added by an indexer at a later date, as a key to the earlier cases. It is an indication of a change in the language. In fact, 'mere puff' became the term for this legal concept in the *Carlill* case. 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. It is, after all, a name for a previously unnamed concept.

6. Conclusion

This demonstration of retrieval has focussed on pattern matching in simple, but realistic, situations for clarity and simplicity. The research includes questions derived from law cases citing those represented in the knowledge base. Greater detail is given about inference in the type hierarchy using the more complex problems as tests. In both descriptions, pattern matching is done among the conceptual representations supplemented by the lexicon and the additional semantic constraints.

---

[7] 'Reward' here is actually a compensation, but is indeed called 're-ward' in the case report.