CHAPTER 5

UNDERSTANDING AND REPRESENTING LEGAL BRIEFS

5.1 INTRODUCTION

The effective problem solving depends on the internal representation of the problem which in turn is heavily dependent on the problem understanding capability of the problem solver. The problem understanding has been recognized as the first stage of problem solving [80]. The depth of understanding and hence the grade of excellence of the internal representation of the problem on hand decisively depends upon the understander's domain knowledge and the experience. Variations in either domain knowledge or experience can materially affect the contents of a problem representation. This naturally has a direct bearing on the solution, since it is clear that appropriate representation of problems is crucial to their ultimate solution [81]. In general, problem understanding is a process that receives an initial problem description from the environment and constructs an internal representation of the problem. The initial problem description can be in any modality. The descriptions are often incomplete. A problem solver always depends on some interface regardless of the fact that the input is in either aural, tactile, visual or textual form. When a problem description is confined to textual form only, then a natural language component will be a significant part of problem understanding. In most of the earlier works on computer based systems the input to the problem solver will be either available readily on the data base or acquired during the consultation process. In either case the data will be in
the form of either associative triples or quadruples ([3], [82]) or as the values of variables representing certain prototypical situations [83] or as a piece of representation indicating only the events (actions and their results) in the order of their occurrence ([14], [84]). For example, in all EMYCIN systems [85] the syntax for statements of fact is:

\[
<\text{attribute}> <\text{object}> <\text{value}> <\text{certainty}>.
\]

In this chapter the problem understanding phase of a legal consultation system that takes a legal brief as its input, interprets and understands it from a legal viewpoint has been discussed. The output of this phase will be an internal representation of the problem that could be readily used by the case analysis or consultative advice stage of the system.

5.2 LEGAL READING

Legal briefs are read very differently by those with legal training than by those without. In order to understand what one has read, one must be able to recognize what is significant in the text, and what is trivial. This process of recognition involves reference by the reader to facts other than the text itself. The significance of the text varies with the factors to which it is connected and will vary with the type of reader and his or her background knowledge.

For example, consider the text given in figure 5.1 above. Different renderers will connect different elements of the text to different external pieces of information. A transport company person reading the above mentioned text understands that the Standard Packagings Ltd., Madras manufactures and supplies the cardboard boxes and feels other information present in the text as irrelevant to him. A legal reader on the other hand identifies that there is a case of
industrial dispute (a general legal context [figure 5.2]) in which the disputed problem is pertaining to the retrenchment of an employee in general (a specific legal context [figure 5.2]) and the effected employee is claiming for the retrenchment compensation (a sub-specific legal context [figure 5.2]) and is not bothered about what the establishment manufactures. After ascertaining the specific and sub-specific legal contexts, the facts relevant only to these contexts are yanked out of the text. This is to say that not all the facts that seem to have legal pertinence

T L Rajamani was appointed by the Standard Packaging Ltd., Madras, on 1st February, 1982 as an accounts assistant in its card board box manufacturing department. He was retrenched from the service on 8th January, 1985. Neither the required notice nor the compensation was given to him. Following the orders made by the Industrial Tribunal he was taken back into the service on 22nd March, 1986. Is the retrenched employee eligible for the retrenchment compensation ?

Figure 5.1 A sample facts situation.

globally, i.e., over an entire legal sub domain, are actually relevant to the current context, i.e., a sub-specific context within a specific context. For example, in the text given in figure 5.1 the current legal context is the sub-specific context of "eligibility for retrenchment compensation" within the specific context of "retrenchment and lay-off". Though the fact that "the employee was taken back into the service at a latter date" may have legal pertinence under some other legal context, however, it is not at all a relevant fact in the current context of deciding about the "eligibility for retrenchment compensation" and calculating the amount payable. Thus the doctrine that is followed by a legal practitioner while reading and understanding a legal brief or report is to
FIG. 5.2 SPECIFIC AND SUB-SPECIFIC LEGAL CONTEXTS
first identify the legal context (i.e., the problem type), call in all the facts (i.e., the legal variables) that are required by utilizing his or her past professional experience, educational knowledge as well as the commonsense knowledge and then to extract and assign the values to these legal variables. It should be noted carefully that every aspect of reading and understanding a legal brief expects the "reader", may it be a program or a human being, to have a prior knowledge of the domain without which "it" can neither identify the contexts (i.e., classify the problem) nor process it further.

5.2.1 The legal variables

In the legal realm, unlike in the domains of medicine, chemistry etc., the data to be interpreted and analyzed will be mostly available in legal briefs which will be in the form of a text. The legal briefs are inherently written using phrases like "the period of continuous service", "there was a strike", "date of closure of an establishment", "the closure was legal" etc. Depending upon the fact situations or cases lawyers "construe" and "modify" the meaning of such phrases. Hence these phrases, with different fact situations, will have different values associated with them and are therefore referred to as legal variables. Construing such legal variables in order to bind them with certain values thus entangles the processes of legal reasoning with the natural language aspect. However, the deeper reason for such an entanglement of the natural language aspect in the legal analysis is due to the open texture [21] of many legal variables. It applies equally to legal sub-domains such as assault and battery [48], corporate taxation [86], manufacturers product liability [87], contract law [14] as well as industrial disputes. The task of problem understanding is to ascertain the relevant legal context, to summon only the needed legal variables and then to assign
Figure 5.3 Representing the legal variables

Proper values for these legal variables. All the legal variables, for this purpose, have been classified into two groups called CIV and CDV. The former group consists of all those variables which are common to all legal sub-specific contexts (i.e., which are current context independent) whereas the latter group consists of those variables which are current context dependent. For example, the values of the variables represented by the phrases like "the appointment date" , "the employer" , "the employee" etc., are required under all legal sub-specific contexts and are, therefore, treated as CIVs whereas the details of "the reappointment of a retrenched employee" are not required in the sub-specific context of "dispute regarding retrenchment compensation" and
Figure 5.4 Service interruption details.
as such the variables associated with the "reappointment" event like reappointment-date, conditions on which reappointment is made, etc., are therefore treated as CDVs.

A variable belonging to either of the groups mentioned above is represented by a slot, having a format shown in figure 5.3, the slot name being the name of the legal variable it is representing and having a "value facet" that grasps the value of the corresponding variable. Some such slots will also have an additional facet called "default value facet" using which default values for them may be assigned. Apart from the above classification, the legal variables have also been classified into two other groups. In this classification the first group consists of those variables the value of which are definitive and the second group consists of those variables the value of which are boolean (i.e. YES/NO type). For example, the variables "strike begin date", "closure date", etc., belong to the former group where as the variables "there was a strike", "the strike was legal", etc., belong to the latter group. The variables belonging to the latter group i.e., to the YES/NO group will have default values associated with them always, (see figure 5.4) where as the former group may or may not have such default values associated with them. For example, the definitive variable "number of strikes" can have a default value of 1 where as "the appointment date" or "the date of closure" can not have any such default values associated with them. Generally, providing a value to a YES/NO type variable brings in more variables for values to be assigned to them. For example, an "Yes" value for the variable "there was a strike" brings the variables like "number of strikes", "strike begin date", "strike end date" etc., (figure 5.4).
5.2.2 Intentions

The significance of intentions associated with the actions performed by the parties involved in any event (like the employer and employee) varies with particular legal area. In Criminal Law intentions are of paramount importance. Also, it is well known that Assault and Battery are intentional torts. However, from the following sentences of Williston [88] it is obvious that in the case of Contract Law the secret intent of actions are inconsequential.

In some branches of the law, especially in the Criminal Law, a person's secret intent is important. In the formation of contracts it was long ago settled that secret intent was immaterial, only overt acts being considered in the determination of such mutual assent as that branch of the law requires.

Whether there exists a contract expressed in writing, in a deed, by word of mouth or by conduct, the general principles of the contract law apply to the industrial employment and hence to the realm of industrial disputes [89]. As such while analyzing the textual form of data, in a legal sub domain like industrial disputes, it is not necessary to meditate the 'intentions'. The omission of handling the intentions has simplified the development of the problem understander.

5.2.3 Commonsense and Default information

As already mentioned, due to the non availability of the information in the text, often, some of the required slots will be left unassigned. This warrants the system to query the user. For example, the data given in figure 5.1 does not mention any thing about the "the total number of workmen employed" in the establishment. Though such an information seems to be trivial, it is very important as it has an effect
on the applicability or non-applicability of certain sections of the ACT. Hence the system asks the user for such data as it was not present in the given textual data. Further, the system exhibits certain amount of general or commonsense knowledge on some occasions. For example, the information that an employee works "above the ground level" or "below the ground level" is crucial in computing the "period of continuous service" which in turn is required to decide whether an employee is "eligible for retrenchment compensation". This information is obtained by considering the "nature of work" the employee carries out in his or her day to day work. In the example textual data, given in figure 5.1, it is said that the "post" of the employee is "accounts assistant" using which the system concludes that the employee works "above the ground level", as nowhere accounts assistants are expected to work below the ground level, thus exhibiting a bit of commonsense.

As indicated already in the previous section the system is capable of assuming default values for the YES/NO type of variables whenever specific values for such variables are not available. However, before actually assuming such default values, the system asks the user whether it can make such assumptions or not. This querying of the user is necessary as it is required to keep the user informed of the assumptions or the inferences drawn based on the commonsense knowledge. At this point the user can respond by saying YES, NO or DONOTKNOW. A DONOTKNOW answer makes the system to generate multiple answers by assuming a YES as well as a NO, one after the other. Obviously, the facility of typing in a DONOTKNOW answer aids in circumventing the situation that may arise when the user is not sure of a definite answer. Particularly such situations will arise whenever hard legal cases are being dealt with.

The assigning of values for all the legal variables
(i.e., slots), is done by a problem understanding program called the SIFTER. The sifter, after binding all the needed variables, proceeds further by generating the facts both in their full form as well as terse form. This output of the sifter will be taken as the "input" by the consultative advice stage of the system.

5.3 SIFTER

Unlike other professions, in the legal profession lawyers does not sift all the facts that may lay abstrused in a legal brief at one hand but sift them only as and when the legal contexts demand such information [26]. Understanding of legal briefs for the purpose of legal reasoning is therefore distributed and follows a "demand driven" approach of the sort suggested by Granger et al., [90]. Whether the facts retrieval is done at a single time point or distributed throughout the 'reasoning' process the functions described below are essential to the problem understanding task.

Understanding a legal brief involves a three stage process:

1. Processing of textual form data for the names, dates and time-intervals information as well as noun phrases pre-processing.
2. Goal or problem type identification, and
3. An augmentation process that continually keeps on sifting the facts depending upon the shifting legal sub-contexts.

The net effect of these processes is the construction of an extended as well as an unambiguous problem representation that will be available for the further problem analysis.
5.3.1 Processing of the textual form data

The processing of any textual form data for the purpose of facts retrieval involves dealing with a fairly large amount of syntactic, semantic, pragmatic and world knowledge. The work done in natural language understanding has been only in one direction out of the multidirectional approaches that one can pursue with. For example, Woods [91], Marcus [92] et al., have concentrated on syntactic parsing, Schank [93] has concentrated on semantic representation whereas Perrault, Cohen, Allen et al., ([94], [95]) have dealt with speech acts. In this work the emphasis has been on the overall understanding of the legal briefs available in textual form and the retrieval of the relevant facts.

At this embryonic stage of problem understanding, a syntax analyzer analyses and gives the noun phrases processed output as shown below.

T. L. Rajamani was appointed by the Standard Packaging Ltd., Madras, on 1st February, 1982 as an accounts assistant in its cardboard box manufacturing department.

```
Input
Syntax analyzer
Output
(NP-1 WAS APPOINTED BY NP-2 ON NP-3 AS NP-4 IN NP-5)
```
The syntax analyzer has been built using the following modules working in tandem (figure 5.5).

(1) Names-dates-periods processor referred to as NDP-processor

and (2) Noun-phrase processor referred to as NP-processor.

5.3.1.1 NDP-processor

Names handling

To start with the input sentences given to the syntax analyzer will be taken up by the NDP-processor which in turn hands it over to the name-handler first. By utilizing the knowledge that the proper nouns will start with uppercase alphabets the NDP-processor replaces the proper nouns by some pseudo names as shown in figure 5.6(a). It has the ability to recognize the beginning of a sentence and accordingly can discriminate between the articles like a, the etc., time
Periods handling

Normally the sentences involving the time adverbial "during" and the appearance of the prepositions "from" and "to", in that order, and having date information just ahead of them carry certain period information with them. The period-handler utilizes this knowledge during the processing of an input list to it and replaces the period, if present, with pseudo names as is done by the name and date handlers (figure 5.10).

In this case also the relation of correspondence between the pseudo names for the dates and the actual periods will be stored in an association list (figure 5.11).

It is obvious that the parties involved, the dates and the periods associated with the events vary from case to case. As such, while using knowledge-based consultation systems it becomes necessary to update the lexicon as and when a new fact situation is taken up for the analysis. As there is a need of referring to the time point and time interval information from many different perspectives depending upon the contexts there is a need for a uniform way of representing the time point (i.e., date) and time interval (i.e., period) information. By employing certain pseudo names like NAME-1, NAME-2, etc., to take care of the proper nouns, DATE-1, DATE-2, etc., to take care of the time points and PRD-1, PRD-2, etc., to take care of the time intervals, the not so much relished work of frequently updating the lexicon is avoided.
(THE WORKMEN WERE ON STRIKE FROM DATE-2 TO DATE-3)

Input

NDP-processor

Output

(THE WORKMEN WERE ON STRIKE PRD-1)

(DURING THE STRIKE THE EMPLOYEE WAS WORKING TEMPORARILY AT NAME-1)

Input

NDP-processor

Output

(DURING PRD-1 THE EMPLOYEE WAS WORKING TEMPORARILY AT NAME-1)

Figure 5.10 Period processed sentences.

Figure 5.11 Periods-a-list.

The processing of names (proper nouns), dates and periods not only reduces the number of constituent members of any sentence but also brings in certain amount of coherence between the various sentences as the same names, dates and periods referred to at more than one place in the input text.
are replaced by same pseudo names. Such a processing also relieves certain responsibilities of chunking the different words, appearing consecutively in a sentence, constituting a single name, date or period as in the case of "The College of Engineering closed down ..........", ".......was working in the Department of Science and Commerce since ....", ".......... on 14th January, 1989", etc.,.

5.3.1.2 NP-processor

Given the NDP-processed output, the NP-processor does the syntactic analysis and outputs its input list in the noun phrases processed form as shown in figure 5.12.

\[
\begin{align*}
\text{(NAME-1 WAS APPOINTED BY THE NAME-2 ON DATE-1 AS AN ACCOUNTS ASSISTANT IN ITS CARDBOARD BOX MANUFACTURING DEPARTMENT)}
\end{align*}
\]

\[
\text{Input} \quad \text{NP-processor} \quad \text{Output}
\]

\[
\begin{align*}
\text{(NP-1 WAS APPOINTED BY NP-2 ON NP-3 AS NP-4 IN NP-5)}
\end{align*}
\]

**Figure 5.12 NP-processed output.**

The noun phrases processed output shown above is based on the syntactic parse tree shown in figure 5.13. The noun-phrase preprocessor identifies the noun-phrase constituents and then replaces these noun-phrase constituents by a single noun-phrase node. A noun-phrase can consist of a single constituent or a group of constituents. Figure 5.14 gives some of the rules used to formulate a noun-phrase node. When the NP processor detects a single constituent or a group of...
Figure 5.13 The syntactic parse tree.
constituents that satisfy one of the above rules it replaces it by a single node NP-K where K is an integer that will be initially 1 and gets incremented just before a new noun-phrase node is created. When words other than noun-phrases occur they are simply added to the structure after checking whether the sentence is syntactically correct or not.

<table>
<thead>
<tr>
<th>np</th>
<th>proper noun</th>
</tr>
</thead>
<tbody>
<tr>
<td>np</td>
<td>date</td>
</tr>
<tr>
<td>np</td>
<td>noun, noun</td>
</tr>
<tr>
<td>np</td>
<td>pronoun, noun</td>
</tr>
<tr>
<td>np</td>
<td>det, noun</td>
</tr>
<tr>
<td>adjs</td>
<td>adj,...adj</td>
</tr>
<tr>
<td>np</td>
<td>det, adjs, noun</td>
</tr>
<tr>
<td>np</td>
<td>art, adjs, noun</td>
</tr>
<tr>
<td>np</td>
<td>art, noun</td>
</tr>
<tr>
<td>np</td>
<td>name</td>
</tr>
<tr>
<td>np</td>
<td>period</td>
</tr>
</tbody>
</table>

Figure 5.14 Rules to formulate a noun phrase node.

The correspondence between the noun-phrase nodes and what they represent is required and as such will be stored in a data base, in the form of an association list on which the noun-phrase nodes like NP-1, NP-2, etc., will be the "keys" and phrases they represent will be the "datum", as shown in figure 5.15.

Figure 5.15 NP-a-list.

```lisp
(((NP-1 NAME-1)
  (NP-2 THE NAME-2)
  (NP-3 DATE-1)
  (NP-4 AN ACCOUNTS ASSISTANT)
  (NP-5 ITS CARDBOARD MANUFACTURING DEPARTMENT))
```
5.3.2 Goal identification

The important step to solve or analyze a problem, using a computer or otherwise, is to understand it and to construct its internal representation. This in turn needs the identification of the "problem type" or the "goal" of the analysis. A problem type or the "classification" guides the selection and instantiation of the pertinent legal variables and thus provides a "context" to formulate the internal representation of the problem. This step of problem classification and providing the legal variables their values is analogous to the "hypotheses formulation" step of medical consultation systems. The procedure employed, as already discussed in section 5.2, assumes that a problem solver's or analyzer's internal problem representation is heavily dependent on his or her domain knowledge. Identification of the problem type may either follow or precede the retrieval of values for the context independent variables (CIVs). The information regarding the problem type will be available in the noun-phrases of the sentences like "The claim is regarding the retrenchment compensation", "The dispute is regarding the compensation upon closure of the company", "Is the retrenched employee eligible for the retrenchment compensation?" etc.. The problem identification is done by scanning through all the pre-processed sentences for the concept "dispute" which is recognized by the presence of the words like dispute, claim or their synonyms in the processed noun-phrases. However, the information regarding the problem type will be customarily available in the interrogative sentences like "Is the retrenched employee eligible for the retrenchment compensation?", "Are the workmen eligible for the lay-off compensation?", "What is the length of continuous service?", etc..

Civil codes, like the Industrial disputes Act, are chiefly statutory type and are available organized in the form of chapters, sections and parts of the sections. Every
section will have a specific heading which clearly indicate its purpose. Every section, therefore, has been treated as a "context" or "problem type" identifiable by its heading. All 'problem types' have been represented by frame like data-structures, the slots of which are the context dependent legal variables. Once a "problem type" is selected, the slot names representing the legal variables behave in a way similar to distinctly stated questions and acquire values by critically examining all the pre-processed sentences in an order. For example, the retrenchment compensation context will select a frame that will be as shown in figure 5.16.

<table>
<thead>
<tr>
<th>NAME</th>
<th>DISPUTE</th>
</tr>
</thead>
<tbody>
<tr>
<td>REGARDING</td>
<td>(RETRENCHMENT COMPENSATION)</td>
</tr>
<tr>
<td>RETRENCHMENT-DATE</td>
<td>NIX</td>
</tr>
<tr>
<td>EMPLOYEE</td>
<td>NIX</td>
</tr>
<tr>
<td>EMPLOYER</td>
<td>NIX</td>
</tr>
<tr>
<td>ENTRY-DATE</td>
<td>NIX</td>
</tr>
</tbody>
</table>

Figure 5.16 Retrenchment compensation context.

The NIX values indicate that the slots are yet to be filled and are filled by executing certain procedures. Most of the boolean type variables are instantiated by searching for certain verbs or their alternative forms. For example, the presence of any of the verbs retrenched or dismissed will provide an YES answer to the question "was there a retrenchment?". This question itself will be represented as a variable "retrenchment", and will be the name of a slot with a proper default value associated with it. Once such a slot's value facet gets an YES answer, it in turn calls other
slots (i.e., variables) to be filled with details from the text. The information necessary to fill the definitive type of slots are usually available in the noun phrases. For example, the information necessary to fill the slot "entry-date" that represents the fact "the date of appointment" will be available in the noun-phrase node NP-3 (figure 5.13), to fill the slot "post", which is further capable of enlightening on the nature of work the employee is engaged with, will be available in the noun-phrase node NP-4 (figure 5.13), etc.,. However, it should be noted that not all the noun phrases are required to be looked into for providing values to specific variables (i.e., slot names). For example, when it is required to assign a value to the entry-date only noun phrases present in the sentence that engrosses "the appointment" concept or information is considered. Any noun phrases processed sentence that contains verbs, corresponding to the action appointment, like joined, entered or appointed is said to contain "the appointment" information. The required information will be in a noun phrase that appears immediately after the preposition "on", to the right of the action word (i.e., appointed, joined or entered) and will be of the type date. The noun phrase processing thus helps in efficient instantiation of the slots. The noun phrase processor also helps in easy retrieval of temporal information which are available as noun phrases following the prepositions like on, from, to etc., and/or the time adverbials as already discussed. The language used in writing the legal briefs has an established mode and hence the domain of possible sentence structures available in the legal brief is limited which facilitates the filling of every slot easily. This is also because a legal consultant is decided on the information he or she expects and wants from the legal text.
5.3.3 Augmenting the problem representation

As already mentioned the augmentation process is distributed almost over the entire reasoning process. Either due to the incomplete data supplied or as a result of the shifted attention of the problem solver some information which may be required very much may be missing. During augmentation process the missing information is inferred or acquired.

<table>
<thead>
<tr>
<th>NAME</th>
<th>DISPUTE</th>
</tr>
</thead>
<tbody>
<tr>
<td>REGARDING</td>
<td>(RETRENCHMENT COMPENSATION)</td>
</tr>
<tr>
<td>RETRENCHMENT-DATE</td>
<td>8/1/85</td>
</tr>
<tr>
<td>EMPLOYEE</td>
<td>(T L RAJAMANI)</td>
</tr>
<tr>
<td>EMPLOYER</td>
<td>(STANDARD PACKAGING LTD.)</td>
</tr>
<tr>
<td>ENTRY-DATE</td>
<td>1/2/82</td>
</tr>
<tr>
<td>APTCAT *</td>
<td>PERMANENT</td>
</tr>
<tr>
<td>BREAK-IN-SERVICE *</td>
<td>NO</td>
</tr>
</tbody>
</table>

* variables APTCAT (appointment category) and BREAK-IN-SERVICE have been appended during the augmentation process.

Figure 5.17 An augmented context

Once the 'problem type' is selected and its slots are filled up, the reasoning process starts with the aid of the 'domain knowledge' (also represented in the form of frame like structure) associated with the selected problem type. Often, the reasoning process keeps changing its focus on different 'problem types'. For example, while determining the "eligibility for the retrenchment compensation" there arises
the need for "the period of continuous service". This, obviously, shifts the reasoner's attention to the problem type "the period of continuous service" which, therefore, is selected and treated as a sub-problem. All the variables (i.e., slots) associated with this new context are collected and appended to the already existing 'problem representation frame'. These variables are also instantiated in a similar way as explained earlier. Figure 5.17 shows an elaborated and instantiated problem representation. The reasoning, thus, continues till a plausible solution is obtained. At the end, a "complete representation" of the problem will be available, which may be used to examine all the legal variables and their values that took part in the reasoning process as well as to produce the summary of the entire reasoning process.

5.4 SUMMARY

The grade of excellence of the result obtainable from any knowledge processing system crucially rests not only on the attributes of the knowledge representation technique that is employed to represent the domain knowledge, the depth of the knowledge the system possesses with it and the control strategy that is employed but also, equally importantly, depends on the system's "problem understanding" capability. This chapter has stressed the need of the domain knowledge as a priori constraint in reading, classifying, understanding and hence constructing an initial representation of a given domain relevant problem which, at present, is being used by the consultative advice stage of a legal consultation system called TIDA [18]. Like medical reports, legal briefs are also written using typical word combinations that are domain dependent. A discussion on the details of the problem understanding program, SIFTER, has also been presented in this chapter. The SIFTER takes care of the natural language aspects in a broader sense by processing the names, dates and periods information and generates noun phrases processed
sentences. The sifting of the relevant facts from these noun phrases processed sentences takes place automatically and, at times, utilizes users' contextual knowledge. The sifting of facts does not take place at a time, but takes place almost throughout the entire period of the consultation process depending upon the "focus-of-attention" of the problem solving system.