# SANDIA REPORT

SAND96–2337 • UC–705
Unlimited Release
Printed September 1996

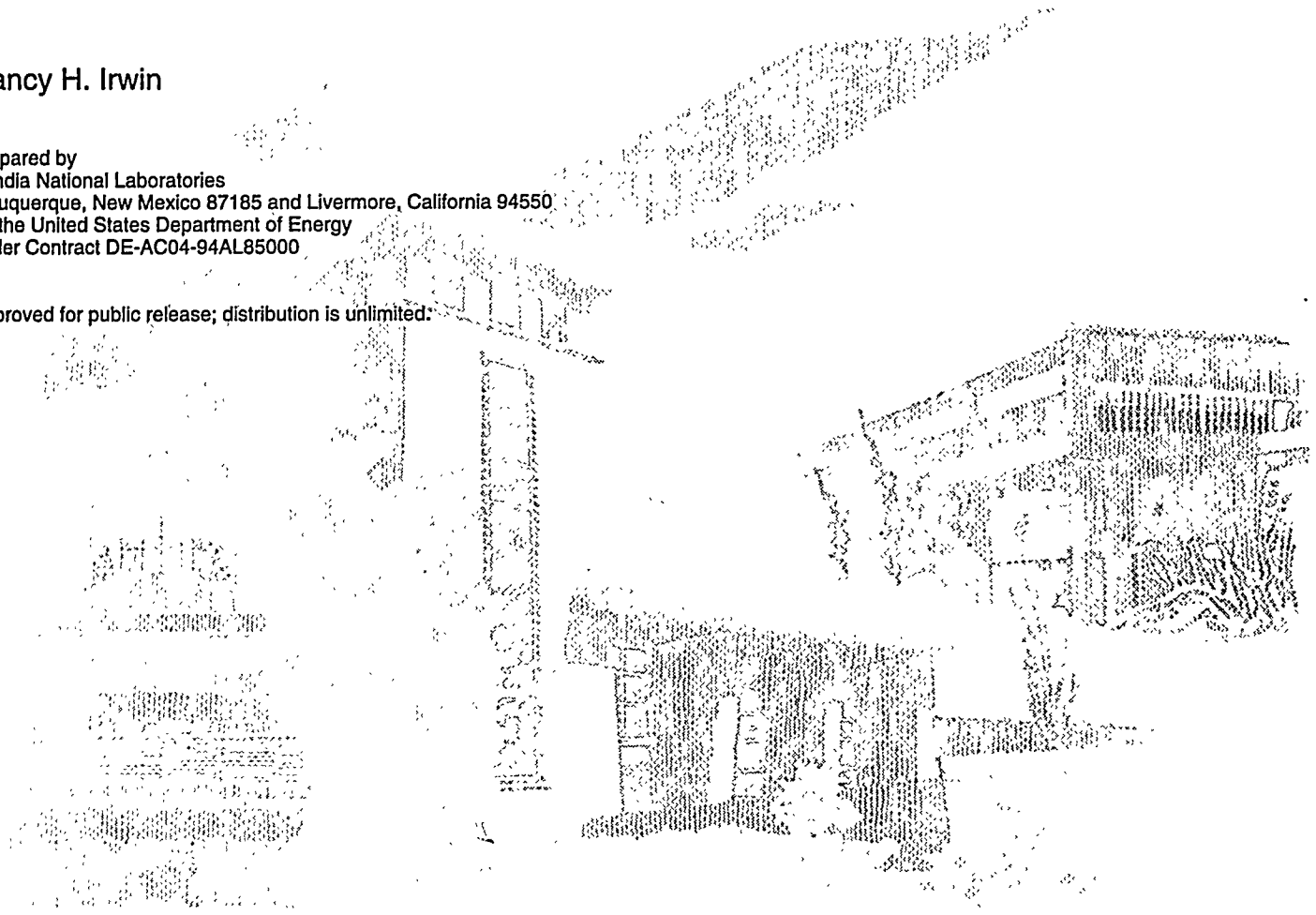
# Domain-Independent Information Extraction In Unstructured Text



Nancy H. Irwin

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550
for the United States Department of Energy
under Contract DE-AC04-94AL85000





SF2900Q(8-81)

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.



Available to DOE and DOE contractors from
Office of Scientific and Technical Information
PO Box 62
Oak Ridge, TN 37831

Prices available from (615) 576-8401, FTS 626-8401

Available to the public from
National Technical Information Service
US Department of Commerce
5285 Port Royal Rd
Springfield, VA 22161

NTIS price codes
Printed copy: A03
Microfiche copy: A01

SAND96-2337
Unlimited Release
Printed September 1996

Distribution
Category UC-705

# Domain-Independent Information Extraction In Unstructured Text

Nancy H. Irwin
Software Surety Department

Sandia National Laboratories
Albuquerque, NM 87185

## Abstract

Extracting information from unstructured text has become an important research area in recent years due to the large amount of text now electronically available. This status report describes the findings and work done during the second year of a two-year LDRD. Building on the first-year's work of identifying important entities, this report details techniques used to group words into semantic categories and to output templates containing selective document content. Using word profiles and category clusterings derived during a training run, the time-consuming knowledge-building task can be avoided. Though the output still lacks in completeness when compared to systems with domain-specific knowledge bases, the results do look promising. The two approaches are compatible and could complement each other within the same system. Domain-independent approaches retain appeal as a system that adapts and learns will soon outpace a system with any amount of a priori knowledge.

# Acknowledgements

This work was funded by the Laboratory Directed Research and Development (LDRD) Program. The author wishes to thank John Taylor, chairman of the LDRD Office's National Security Technology Program Area, Sharon Chapa, Manager of the Software Surety Department, and Sharon DeLand for their enthusiastic support. Thanks also to Kim Culp M.D. and the Scott and White Hospital in Temple, TX for the use of sample medical data.

# Contents



## Figures

# Domain-Independent Information Extraction In Unstructured Text

## Background

One long-range vision of information extraction is a system which could gobble up lots of text and then allow a user to sit down at a terminal and ask questions in natural language that the system would then answer in an intelligent way. This is still considered an ambitious task in spite of the many years already invested in this area.

A major component of a natural language system provides the means for extracting and modeling the information in the document. The proposed approach to this activity was based on an entity-relationship concept similar to that used in E-R models. The entities were compared to nouns while the relationships were compared to verbs. The first year's milestones centered around identifying the entities and met with reasonable success. The results of this first year's work are documented in last year's report [IRW95].

Briefly, entities are identified in a system that builds word tokens, assigns parts of speech, does minimal noun phrase parsing and obtains semantic meanings from a public domain, semantic net data file [WOR93]. Important entities and compound noun phrases, such as "cruise missile" and "heart attack", are identified largely through profiling. Unknown words are analyzed using a pattern matcher approach plus a few heuristics for proper names.

Using these entities as a reasonable approximation of the set of important words in the document, the work of the second year has focussed on the relationships between those words and the identification of meaningful subtopic groupings.

Sets of documents from three domains have been in use during the development process: treaties, medical data, and some baseball newspaper shorts. Later work has focused more on the large set of aliased doctor/patient transcribed sessions made available to us by Scott and White Memorial Hospital in Texas.

## Current Goals

The original proposal was clear is stating that the interest was in exploring techniques that were domain independent. Current research systems target a domain and then build knowledge specific to that domain. Since this is a time-consuming process, recent effort has investigated how to build the domain knowledge quicker and more automatically. This LDRD proposed finding domain-independent approaches that would not require pre-defined domain-specific knowledge. Such a system would then be able to handle text outside of the expected domain. The approaches discussed here could be used in a system which combines approaches from both camps.

Since information extraction is a large problem, the following bullets clarify what has been considered within or not within the scope of investigation of this LDRD.

- The system will not try to generate an answer to a query. It may locate in the document where a reference can be found or it will output frames of the information found in the document. This eliminates the need for full text regeneration and relieves the system of the correctness of

response responsibility associated with the subtleties of a natural language answer.

- The proximity relationship between entities will be preserved. In one demonstration proximity is defined as in the same sentence; the second uses the medical record as the boundary. Proximity could also be defined as the same paragraph or the same section as in a textbook.
- Instead of trying to model all the information in a document, a training run evaluates the document content. A training run looks at a sample document, determines the subcategories of high intersection amongst entities and favors additional entities in those subcategories. In the medical domain, for example, if the training document contained entities whose subcategories clustered around drugs, symptoms, and body parts, then nouns in those categories would be sought out as further additions to the entity set. Thus one important relationship in the E-R concept will be the membership of a noun in one of the selected semantic categories.

The last bullet above is the area in which this project's second year has mostly focussed. A non-interactive training session identifies entities and looks for clusters of subcategories (amongst the entities) that can later be viewed and edited if desired. These categories and subcategories direct the system in favoring further information to be extracted. The public domain semantic dataset being used in this implementation is WordNet. The categories and names used in WordNet will be reflected in the output of both the training session and later the interactive system.

## Previous Work

In 1992 the DARPA-sponsored Fourth Message Understanding Conference (MUC-4) focused on a competition of participants' text analysis systems. The task was to extract information in the form of templates from newswire articles on terrorism in Latin America. Consecutive years' conferences have continued this competition environment by selecting a new domain each year.

Information in the defined template included items such as perpetrator, victim, instrument, location, date, etc. Participants of the experiment found that it was time-consuming to build the required knowledge. Words such as 'bomb', 'murder' and 'attack' were used as keys to trigger frames and patterns for extracting the desired information. Building tools to speed up this knowledge building process has thus become a secondary focus of information extraction. AutoSlog[RIL93] from the University of Massachusetts and PALKA[KIM93] from the University of Southern California are two examples.

Natural language researchers have also become interested in concept databases like WordNet[WOR93]. Words are connected to other words in hierachies organized through "is-a" and "part-of" relationships. With access to this data, resolution of word sense in a particular context is still necessary. These databases do not provide the role-mapping that some of the domain-specific knowledge bases provide. For example, the actor of a 'buy' is the recipient, while the indirect object of a 'sell' is the recipient. Just knowing that they are both transfer events does not provide the same depth of information. This is information that would enrich this system as well as the domain specific ones that have coded this information for a limited word set.

## Comparison of Output

In a MUC system (selectively extracted from [KRU94]), one might expect output that looks something like this:

**Input**

San Salvador, 19 Apr 89 -- San Salvadoran President-Elect Alfredo Christini condemned the terrorist killing of Attorney General Roberto Garcia Alvarado and accused the Farabundo Marti National Liberation Front of the crime.

**Output**

| | |
|---|---|
| INCIDENT: LOCATION | San Salvador |
| INCIDENT: DATE | 19 Apr 89 |
| INCIDENT: STATE OF EXECUTION | accomplished |
| INCIDENT: INSTRUMENT TYPE | bomb |
| PERP: INCIDENT CATEGORY | terrorist act |
| PERP: ORGANIZATIONAL ID | "Farabundo Marti National Liberation Front" |
| PERP: CONFIDENCE | accused by authorities |
| HUM TARGET: NAME | Roberto Garcia Alvarado |
| HUM TARGET: DESCRIPTION | Attorney General |

The output of our system looks like this:

**Input**

Hamish, Korry # 42410033 EMERGENCY ROOM NOTE 9/4/91
The patient is a 30 year old white male with sore throat for approximately 2 months. The patient had some neck swelling last week that he claimed was painful to touch. At that time he took some Penicillin pills that he had, approximately 6 tablets, which made the swelling go away but the sore throat remained. He complains of loss of voice in he early a.m. non productive cough.

**Output**

| | |
|---|---|
| PATIENT NAME: | Hamish, Korry |
| PATIENT NUMBER: | 42410033 |
| INFO SOURCE: | EMERGENCY ROOM NOTE |
| DATE: | 9-4-91 |
| AGE/GENDER: | a 30 year old white male |
| ILL_HEALTH/AFFLICTION: | sore_throat |
| SYMPTOM: | some neck swelling |
| DRUG/SUBSTANCE: | some Penicillin pills |
| DRUG/SUBSTANCE: | 6 tablets |
| ILL_HEALTH/AFFLICTION: | the sore_throat |
| SYMPTOM: | early a.m. non productive cough |

Though the input to the two is different, this example is included to point out a few similarities and differences. In the first, the informational nugget types to be extracted are predefined. The designer of the system decides ahead of time that a *perpetrator* role will exist with subroles including *incident category*, *organizational id*, and *confidence*. The subroles will have associated with them constraints, patterns and rules to guide the filling from the text. Some constraints will likely include IS-A restrictions which map to a conceptual database. For example, the filler of the role "incident: instrument type" may include a constraint that requires it to be either a *weapon* or an *electronic_device* and since *bomb* "is-a" *electronic_device* it would be a valid filler.

The proposed system also depends on IS-A relationships contained in a concept database, in fact even more so. What nuggets are to be extracted are not predefined by hand but rather are determined by a pre-process profiling pass [IRW95]. This pass can be done on a document by document basis or in the case of many documents in a similar domain, a training set could be used. In the medical records sample, a training set run resulted in the clustering of entities (nouns)

around semantic categories, *symptom*, *ill_health*, *body_part*, *drug*, and *substance*. In this environment, the initial set of semantic category names serves as both the name of the roles and its constraint for fillers. After the training run, there is opportunity for human intervention if desired. Semantic categories can be combined with others; categories can be added or deleted; specialization or generalization of categories can be refined.

Because the information to be extracted is decided upon in an automatic way, what information is to be extracted could be changed on the fly. In a system where the frame definition depends wholly on human activity beforehand, encountering a document in a new and unexpected topic area results in either nothing being extracted or unexpected nuggets filling the pre-defined templates.

The terrorist article has a by-line containing the location and date of the report. The medical record has a header line containing the patient name, medical record, clinical source and date. Both can take advantage of the expected format without using sophisticated processing.

## Clustering of Entities

Finding the correct level of specificity within the hierachy is important as it is used both as a name to the category and as a constraint for what nouns will be selected and how they will be grouped. If the category is too general, accepted fillers will contain many errors. If the category is too specific, good fillers will be missed. Figure 1 shows a part of the conceptual database as relates to *pain*, *headache*, *fever* and *cough*. Most of the words have several meanings but there is a clustering of the words under the concept *symptom*. Selecting the concept *information* would be too general; selecting the concept *pain* would be too specific; selecting concepts *act* or *feeling* would be the wrong sense for the medical records document set. How the entities cluster determine what categories will be selected and then what information will be extracted for the document.

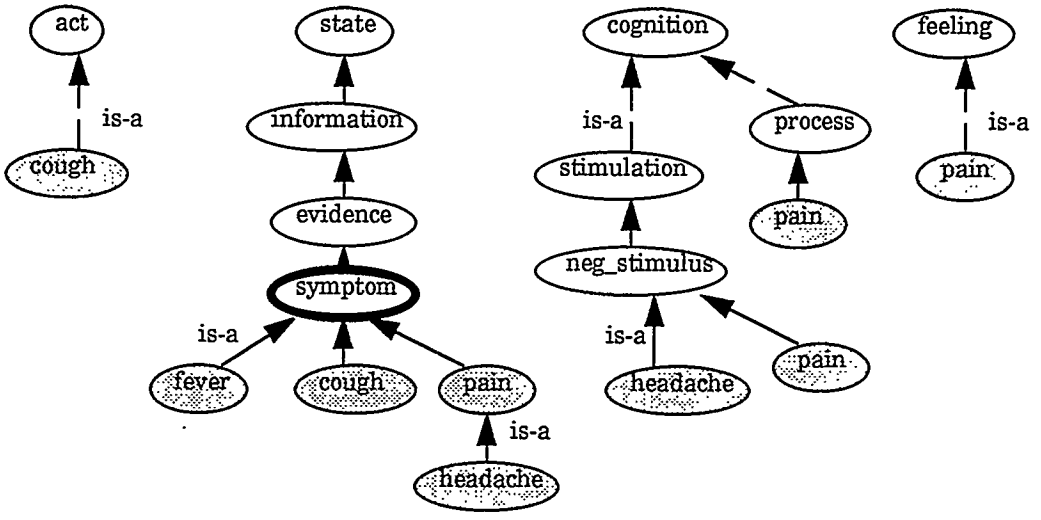



**Figure 1: Concept entries as relates to fever, cough, pain and headache.**

Figure 2 shows four intersecting categories *(drug, symptom, headache, and pain)* for nine entities *(drug, headache, migraine, morphine, narcotic, nausea, pain, redness, tenderness)* in the medical domain. Where *headache* is a subcategory of *pain* and then *pain* is a subcategory of *symptom*,

selecting all 3 categories even if the count were high on all three would defeat the grouping process. So in addition to considering the counts on intersecting categories, it is also important to minimize cases where parent and child categories are both selected. In this case, it would be more intuitive to select the most general category *symptom* and in fact the count supports this selection in this case. In some cases, however, intuitively one would want to select the more specific. If two have exactly the same count, the more specific category is more often than not the correct one. No formula based solely on count and specificity was found, however, that reliably gives the intuitive selection. Favoring the more specific picks up less errors and favors the category of more knowledge and has been the trend. As stated earlier, the user also has the option of changing the selection after the training run.

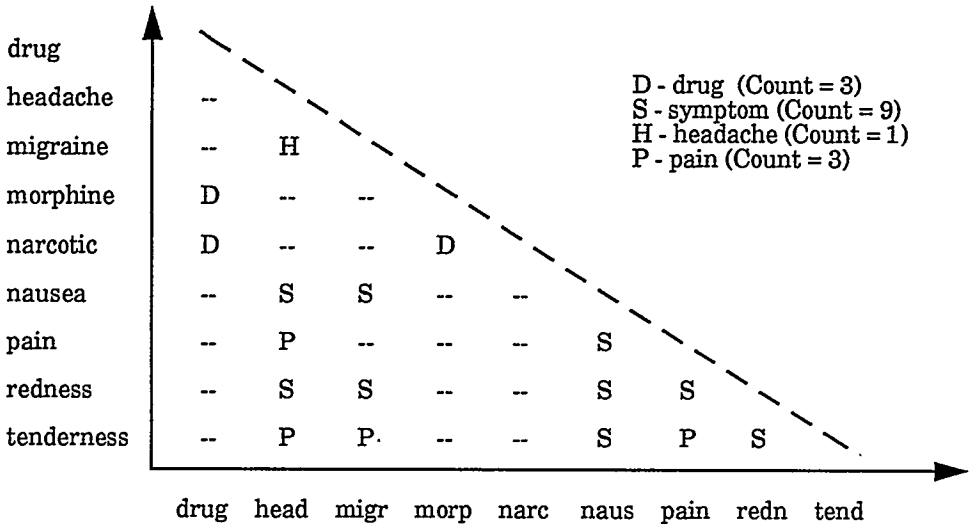



**Figure 2: Category intersections of nine entities.**

Implementation of a more efficient clustering algorithm would facilitate using larger documents for the training run and open the possibility for continuous background updating and learning during normal processing.

## Verbs As Relationships

Earlier this year several attempts were made to force the verbs (as proposed) into the relationship role in the E-R model concept. This has met with several problems.

- Verbs have many semantic senses. Deciding which semantic sense applies in a particular context is more difficult for verbs than it was for nouns.
- Verb senses are of a general nature providing the glue of the sentence. Matching and combining information against these general terms has not been fruitful.
- How to combine across sentences using verb triggers without domain knowledge and domain scripts is still not clear.
- A model of the document information that provides full content availability (not script driven) may be impractical due to the resulting size.
- Getting a correct full sentence parse with full attachment of subjects, objects, and direct objects to verbs and prepositional phrases to appropriate nouns and then correctly interpreting what those attachments mean semantically is ambitious.

Having abandoned this path, more effort went into exploring the semantic grouping of the entities.

## System Description

The implemented approach uses a two-pass process as shown in Figure 3. During pass 1, words are profiled for number of occurrences and for co-occurrence with other words. Nominal compounds, important keywords, and some unknown words are identified and preprocessed. Still in pass 1, the set of frequently occurring nouns are then analyzed for clusters within semantic subcategories. By the end of pass 1, the preliminary set of topic subcategories has been selected. This list can be modified by hand, if desired, before starting pass 2. Results of the first pass provide input into the processing of pass 2. During pass 2, identification of the entities with membership into the selected subcategories is targeted. The output can then take the form of either templates as in Appendix A or tables that can be read by an interactive search engine program as in Appendix B.

In both passes, text documents are processed on a sentence by sentence basis where a tokenizer divides sentences into word and punctuation tokens. Parts of speech are assigned to all words and semantic categories to nouns. In pass 2, partial parsing disambiguates for part of speech and builds the noun and verb phrases. Also by pass 2, the set of possible semantic categories has been limited to those subcategories selected in pass 1.

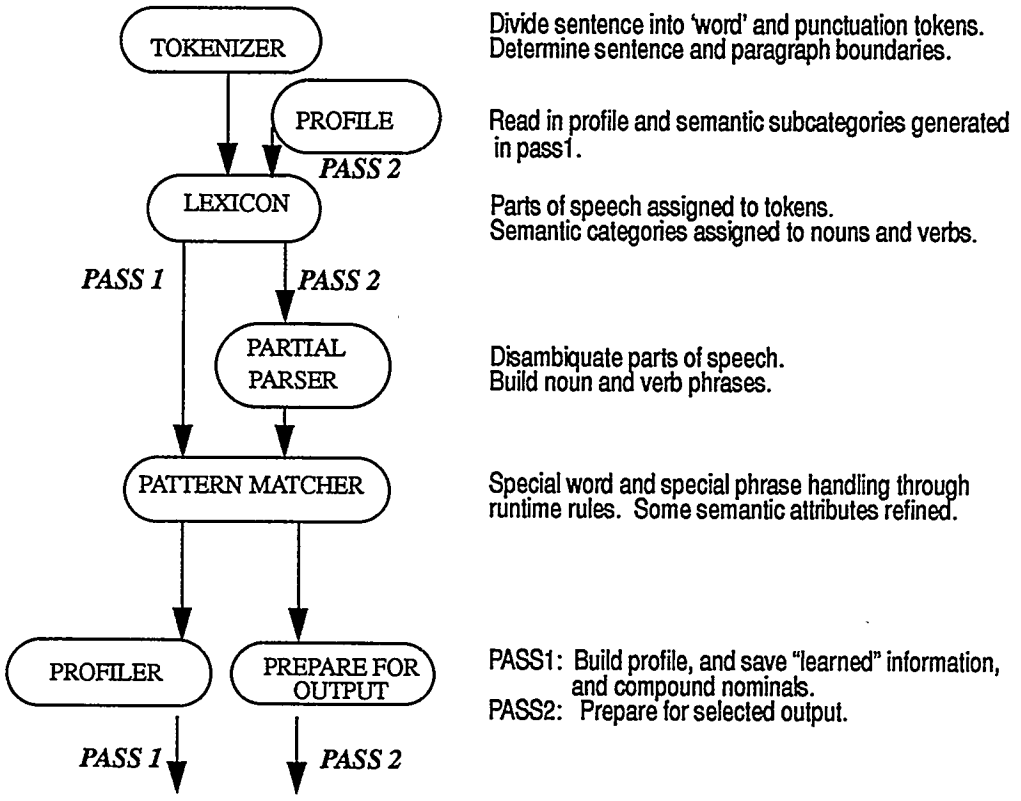



**Figure 3: Block Diagram of the Information Extraction System**

The parts of speech assignment is accomplished using several sources.

- A "special word" word list handles frequently used words and words needing particular attributes not supplied by other sources.
- PC KIMMO, a public domain lexicon, provides basic parts of speech such as adjective, singular noun, adverb, etc. for a larger set of words.
- Suffix analysis provides basic parts of speech to unknown words. E.g., -ly implies adverb; -ing implies present participle. Suffix analysis is also used to get the root form of a word which is then used in coordination with other approaches.
- Existence as a noun in the semantic net provides part of speech *noun* to otherwise unknown words.

The semantic category assignment of nouns and verbs depends heavily on the data supplied with WordNet, a public domain linguistic database. Among other things, WordNet groups nouns into the following categories: act, animal, artifact, attribute, body, cognition, communication, event, feeling, food, group, location, motive, object, person, phenomenon, plant, possession, process, quantity, relation, shape, state, substance, and time. Words with multiple senses often have multiple categories. To this original set of categories was added amount, date, reference, and acronym. About 7500 first names were also added to the data.

WordNet also supplies categories for verbs: body, change, cognition, communication, compete, consume, contact, creation, emotion, motion, perception, possession, social, stative, weather and process. Since half the nouns in the English language are also verbs and several of the categories overlapped already (cognition, communication), processing was simplified by folding all the verb categories into the noun categories. *Emotion*, for example, was mapped to *feeling* and *stative* mapped to *state*. Also from the point of view of understanding language content, it is probably not important as to whether the sentence used the noun or verb form ("made a purchase" vs. "purchased"). As explained in the previous section, however, verb semantic attributes were not pursued in the later experiments.

This section is similar to last year's report [IRW95] as there has been little change to the overall system design. Added however is the emphasis on subcategory grouping. The final output also looks quite different. The emphasis on noun attachment to verbs has been nearly eliminated, while more work has been shifted to the pattern matcher because of the meaning context it can provide to noun entities. With the reduced dependence on verbs as the inter-sentence unifier, the semantic grouping of the nouns has taken on some of the roles originally expected to have been filled by the verb semantics.

## Building Patterns

The pattern matcher has taken on a stronger role than last year. The matcher provides facility for the extraction of information and assignment to categories in a fast and easy to program fashion. Although the rules are written in a domain specific way, it is thought that this process may be automatable by looking at the words surrounding those selected during clustering.

Drugs, for example, are a class of words of interest in the patient medical record domain. The list of drugs however is continually growing. Having identified drugs as important information to extract, patterns known to contain drugs can enrich that extraction. A semantically unknown word, *Z*, or a word *Z* containing *substance* amongst its multiple senses could be assigned the category *drug* if the word appeared in a particular pattern in the phrase dictionary. Drug-related patterns might include the following: <number> mg of *Z*, tablets of *Z*, prescription for *Z*, *Z* for NOUN_disease.

Many statistical language models use a single layer of word class such as part of speech. The pattern matcher however allows for multiple layers of classification. This idea, though applied

differently, is similar in concept to that proposed by McMahon and Smith [MCM96]. Consider the following sentence:

(1) The boys ate the sandwiches.

This sentence would match any of the following patterns:

| | | | | |
|---|---|---|---|---|
| (2)<determiner> | <noun> | <verb> | <determiner> | <noun> |
| (3)<determiner> | <noun_plural> | <verb_trans> | <determiner> | <noun> |
| (4)[opt_det] | <noun_animate> | <verb_animate> | [opt_det] | <noun_inanimate> |
| (5)[opt_det] | <noun_human> | <verb_consume> | [opt_det] | <noun_food> |
| (6) | <lexical_boys> | <lexical_ate> | <lexical_the> | <lex_sandwiches> |
| (7) | <noun_human> | <lexical_ate> | [opt_det] | <noun_inanimate> |

Patterns 2 and 3 are strictly based on parts of speech and syntax. Patterns 4 and 5 depend on semantic information. Pattern 6 requires an exact match to the words. Pattern 7 uses a combination of the others. The pattern matcher as implemented allows use of a combination but does require at least one lexical. This restriction allowed for fast caching and retrieval of patterns.

What is interesting about McMahon and Smith's study is that their goal was to build the word groupings hierachy by using statistical methods on word occurrence clustering in large corpuses. That is, "given a stream of words N, what is the probability that the next word will be the word W". Two words with high probability against the same stream N would cause them to be highly coupled in the hierachy. If successful, their hierachy could replace WordNet and PC-Kimmo in our system. If a semantic hierachy could be built directly from the corpuses, then the next step of automatically building patterns for the pattern matcher appears obtainable. In fact, even if they can't fully build the hierachy, the fact that initial results look interesting, lends credence to the hope that patterns could also be derived in a similar fashion.

One thing that's troubling about WordNet is that it is not totally consistent. *Coffee*, for example, is a *beverage*, while *soda* is a *drink*. Their hierachies only intersect at *food*. Building patterns that depend on word stream occurrence in the language and then using a hierachy built on the same principles would appear to avoid these inconsistencies.

In one composite file of 15 patient encounters, the following were identified as drugs:
*acid, alcohol, amoxicillin, cigarette, cocaine, darvocet, ecstasy, grass, lithobid, marijuana, motrin, prozac, repetab, seconal, smoke, speed, tegretol, tylenol, upper*, vicodin.*

(*: *Upper* is erroneously marked as a *drug* due to its *narcotic* noun meaning. )

These words occurred in the original document in the following context:

| | | |
|---|---|---|
| not take her | [DRUG:*Prozac* ] | on a consistent |
| to take her | [DRUG:*Prozac* ] | on a regular |
| Will use | [DRUG:*Amoxicillin* ] | and they are |
| to take | [DRUG:*Motrin* ] | or Tylenol |
| take Motrin or | [DRUG:*Tylenol* ] | as needed for |
| to home with | [DRUG:*Seconal* ] | for sleep and |
| for sleep and | [DRUG:*Vicodin* ] | for pain relief |
| Negative for | [DRUG:*alcohol* ] | or drug use |
| She used to | [DRUG:*smoke* ] | prior to pregnancy |
| well on the | [DRUG:*Prozac* ] | |
| being on the | [DRUG:*Prozac* ] | her energy and |
| She will continue | [DRUG:*Prozac* ] | 20 mg |
| | [DRUG:*Darvocet* ] | for pain , |

| | | |
|---|---|---|
| previous medications | [DRUG:*Lithobid* ] | 300 mg |
| need for increased | [DRUG:*Lithobid* ] | dosage as well |
| He has taken | [DRUG:*Motrin* ] | without much relief |
| of the left | [DRUG:*upper* ] | lumbar paraspinous muscles |
| to add Proventil | [DRUG:*Repetab* ] | 4 mg at |
| with drugs and | [DRUG:*alcohol* ] | |
| that he uses | [DRUG:*marijuana* ] | when I |
| that he drops | [DRUG:*acid* ] | every weekend |
| He has used | [DRUG:*Ecstasy* ] | one time , |
| one time , | [DRUG:*Speed* ] | two times and |
| two times and | [DRUG:*Cocaine* ] | six times |
| continued to use | [DRUG:*alcohol* ] | |
| possibility of an | [DRUG:*alcohol* ] | problem in himself |
| She does not | [DRUG:*smoke* ] | , does not |
| sensory exam of | [DRUG:*upper* ] | and lower extremities |
| to shoulder and | [DRUG:*Motrin* ], | 800 mg |
| in the right | [DRUG:*upper* ] | extremity without voluntary |
| p.r.n. | [DRUG:*Tylenol* ] | |
| | [DRUG:*Tylenol* ] | and caffeinated drinks |
| wound was numerous | [DRUG:*grass* ] | and dirt |
| on the left | [DRUG:*upper* ] | rib cage area |
| emergency room with | [DRUG:*Motrin* ] | 800 milligrams to |
| She continues to | [DRUG:*smoke* ] | |
| and placed on | [DRUG:*Amoxicillin* ] | |
| she takes the | [DRUG:*Amoxicillin* ] | she vomits , |
| | [DRUG:*Cigarette* ] | abuse |

Visual examination of the context shows the prescribed drugs frequently occurring in one of the three following patterns: (1) a direct object of the verb *take* or the verb *use*, (2) followed by a *number* and *unit of measure* (eg., 3 mg), and (3) followed by the preposition *for* and then a *symptom* or *physiological_state* (eg., for sleep). Currently, these can be input as patterns, manually-encoded, into the pattern matcher. Automatic detection of the patterns has not yet been accomplished, but would be a reasonable and likely obtainable, but not trivial, next goal.

## Current System Demonstratables

Once the subcategories are defined, the system can go off and build its model of the information. These two steps can be done without human interaction if no editing of subcategories is desired. At the moment this is a rather slow process, but it would be done off-line from the user.

The interactive session response time, on the other hand, is fast. Sentence starting and ending locations have been saved as offset address values into the file. Finding the sentences where words or co-occurrences of words appear, is done with a bit compare. The work associated with grouping words into categories and subcategories has already been done off-line.

The current demonstration under various setups shows the following:

1. Full list of entities
2. WordNet subcategories from clustering of entities
3. Subcategory assignment to entities
4. List of entities in any particular subcategory
5. Sentence numbers in document of entity occurrence
6. Actual sentences where entity or group of entities exist in document

7. Aliasing of document to new name set
8. Frames with roles (category names) and fillers (document nuggets). (see Appendix A.)

## Conclusion and Future Work

The TIPSTER project has funded years of work in the area of information extraction. Research systems have come a long way, but much of the work has not found its way into commercial products. Keyword-driven information retrieval systems have been marketed widely, but their capabilities are much more limited. The work of this LDRD has fallen somewhere in the middle of these two areas. The attempt has been to experiment with domain-independent, and therefore widely applicable, methods with an eye toward retrieving the richer kinds of information currently being extracted in the research systems.

Appendix A shows the output of the current system with an appearance closer to that of Tipster output by using medical session boundaries instead of sentence boundaries and outputting subcategories names and fillers at the end of each section instead of waiting for later querying of the system. Appendix B shows a session that is closer in presentation to search engines. None of the techniques experimented with here are incompatible with the domain-specific techniques and both types of techniques could be used the same system. There is clearly appeal in having a system that adapts and learns as it will soon outpace any amount of a priori knowledge.

Application of these techniques are viable in areas such trend analysis, document searching, information compression and aliasing. As naive users are not in a position to encode knowledge about their domain, statistical methods and knowledge acquisition methods remain attractive. Using WordNet in this system resulted in limitations due to the inconsistency of the hierachies grouped and linked by humans. To achieve the next level of knowledge acquisition by automatically building patterns for the pattern matcher would appear to require a semantic network that it more in tune with detectable linguistic patterns. [MCM96] shows promising initial results in the area of automatically building a word classification system using tagged or untagged corpora. Following this research and applying those concepts to the pattern matcher knowledge acquisition component may be worth pursuing.

Though some of these techniques look interesting and promising, no metrics have been applied against other systems as this technique does not duplicate the output of other systems. The system was built to provide a platform for running experiments with different techniques and is not sufficiently robust for general use, nor has it been run on very large data sets.

## References


[ANT90] Antworth, Evan L., "PC-KIMMO: A Two-level Processor for Morphological Analysis", Summer Institute of Linguistics, Dallas, TX, 1990.

[CHU95] Chung, Minhwa; Moldovan, Dan I. "Parallel Natural Language Processing on a Semantic Network Array Processor", IEEE Transactions on Knowledge and Data Engineering, Vol. 7, No. 3, pp. 391-405, June 1995. (SMU, Dallas, TX)

[IRW95] Irwin, Nancy H.; DeLand, Sharon M.; Crowder, Stephen V. "Extraction of Information from Unstructured Text", Sandia National Laboratories Internal Document number SAND95-2532, Unlimited Release, November 1995.

[KIM95] Kim, Jun-Tae; Moldovan, Dan I. "Acquisition of Linguistic Patterns for Knowledge-Based Information Extraction", IEEE Transactions on Knowledge and Data Engineering, Vol. 7, No. 5, pp. 713-724, October 1995.

[KNI94] Knight, Kevin; Luk, Steve K. "Building a Large-Scale Knowledge Base for Machine Translation", Proceedings of 12th National Conference on Artificial Intelligence (AAAI-94), pp. 773-778, July 1994. (USC/Information Sciences Institute, Marina del Rey, CA).

[KRU94] Krupka, George; Jacobs, Paul; Rau, Lisa; Childs, Lois; Sider, Ira. "GE NLTOOLSET: Description of the System As Used for MUC-4", Proceedings Fourth Message Understanding Conference (MUC-4), pp. 177-185, June 1994.

[LEH94] Lehman, Jill Fain. "Toward the Essential Nature of Statistical Knowledge in Sense Resolution", Proceedings of the 12th National Conference on Artificial Intelligence (AAAI-94), pp. 734-741. July 1994. (CMU, Pittsburgh, PA).

[MCM96] McMahon, John G.; Smith, Francis J. "Improving Statistical Language Model Performance with Automatically Generated Word Hierarchies", Computational Linguistics, Volume 22, Number 2, pp. 217-247. June 1996.

[RAU89] Rau, Lisa F.; Jacobs, Paul S.; Zernik, Uri."Information Extraction and Text Summarization Using Linguistic Knowledge Acquisition", Information Processing and Management, Vol. 25, No. 4, pp. 419-428, 1989. (GE CR&D, Schenectady, NY)

[RIL93] Riloff, Ellen. "Automatically Constructing a Dictionary for Information Extraction Tasks", Proceedings of the 11th National Conference on Artificial Intelligence, pp. 811-816, AAAI Press, Menlo Park, CA, 1993. (UMASS, Amherst, MA)

[RIL94] Riloff, Ellen; Lehnert, Wendy. "Information Extraction as a Basis for High-Precision Text Classification", ACM Transactions on Information Systems, Vol. 12, No. 3, July 1994. pp. 296-333. (UMASS, Amherst, MA)

[WOR93] WordNet 1.4 Copyright 1993, public domain software, Princeton University.

## Appendix A: Template Example

This example uses the same underlying processing as in Appendix B but this one shows the output in a format more similar to the TIPSTER output. While this example shows simply the input and output, Appendix B details the immediate steps. Where the first line of each record contains the patient name, medical record number, clinic source, and date, knowledge of that information is used directly.

### Input

*Meshach,Slugg # 00259084 ER-ORTHOPEDIC NOTE 9-3-91*
*The patient is a 66 year old, white female who slipped on a pebble last night and fell on her outstretched right hand breaking her fall. She noticed some pain and gradually increasing swelling of the right wrist, but there was no significant deformity. She felt that it would resolve on its own. However, the pain intensified and she comes in this morning for evaluation.*

### Output

PATIENT NAME: Meshach, Slugg
PATIENT NUMBER: 00259084
INFO SOURCE: ER-ORTHOPEDIC NOTE
DATE: 9-3-91
AGE/GENDER: a 66 year old , white female
DRUG/SUBSTANCE: a pebble
BODY PART: right hand
SYMPTOM: some pain
SYMPTOM: increasing swelling
BODY PART: the right wrist
ILL_HEALTH/AFFLICTION: no significant deformity
SYMPTOM: the pain

### Input

*Commandos, Brenstein # 00545584 EMERGENCY ROOM NOTE 9/4/91*
*CHIEF COMPLAINT: Motor vehicle accident.*
*HISTORY OF THE PRESENT ILLNESS:*
*This 65 year old white female was the driver of a car which was struck from behind. She is complaining of neck, left sided shoulder, and lower back pain. She denies any chest pain, shortness of breath, diaphoresis, abdominal discomfort, nausea, vomiting. She has had no significant medical problems in the past. She has not seen a physician for years. She is on no medications, had no allergies, and no significant past medical history to contribute.*

### Output

PATIENT NAME: Commandos, Brenstein
PATIENT NUMBER: 00545584
INFO SOURCE: EMERGENCY ROOM NOTE
DATE: 9-4-91
ILL_HEALTH/AFFLICTION: motor_vehicle_accident
AGE/GENDER: 65 year old white female
BODY PART: neck , left

BODY PART: shoulder
SYMPTOM: back pain
BODY PART: breath , diaphoresis
SYMPTOM: abdominal discomfort , nausea , vomiting
DRUG/SUBSTANCE: no medications

## Input

*Grobbelaar, Hired # 00583305 CIM CLINIC NOTE 9-2-91*
*CHIEF COMPLAINT: The patient would like to restart birth control pills.*
*HISTORY OF CHIEF COMPLAINT: The patient is a very pleasant 21-year-old female who comes in for a baseline physical exam. She started with the Health Plan and wanted a refill on her birth control pills and also a regular checkup. She stopped her Triphasil 28-day birth control pills in July of 1991. She just ran out of her prescription and was in the middle of changing doctors and insurance companies. Since then she feels like she might have ovulated during the month but hasn't had any abdominal discomfort since. Her last menstrual cycle started on 8-15-91. They usually last five days. She has no cramps. She complains of no breast pain or urinary symptoms. Her last Pap smear was in June of 1990 and was normal. She has never had an abnormal Pap smear. She has had a small amount of fibrocystic disease in her left breast noted for several years and it hasn't changed.*

## Output

PATIENT NAME: Grobbelaar, Hired
PATIENT NUMBER: 00583305
INFO SOURCE: CIM CLINIC NOTE
DATE: 9-2-91
DRUG/SUBSTANCE: restart birth_control_pills
AGE/GENDER: a very pleasant 21-year-old female
ILL_HEALTH/AFFLICTION: physical exam
DRUG/SUBSTANCE: her birth_control_pills
ILL_HEALTH/AFFLICTION: regular checkup
SYMPTOM: abdominal discomfort
SYMPTOM: no cramps
SYMPTOM: no breast pain
ILL_HEALTH/AFFLICTION: fibrocystic disease
BODY PART: her left breast

## Input

****CLINIC NOTES*** 9/09/91*
*Higher,Oberlander # 00908983 EMERGENCY ROOM NOTE 9-3-91*
*The patient is a 72 year old, white male with a history of recurrent large cell lung carcinoma. The patient was last discharged from the hospital on 8-27-91 where he was treated for recurrent left-sided pleural effusions. The patient was found by pleural biopsy and pleural fluid analysis to have recurrent large cell CA. A pleurodesis was attempted that time with Tetracycline to try to reduce the patient's effusions and to decrease his shortness of breath. The patient had a chest tube placed at that time for the pleurodesis and the tube was removed without difficulty or large residual or without any evidence of significant pneumothorax. The patient's baseline is to be generally dyspneic. He was sent home 8-27-91 with continued follow-up with Dr. Porky and Dr. Hanada. The patient was to return today at 11:00 a.m. for an appointment Dr. Porky to look at the chest tube site and determine if any further treatment was necessary. The patient's wife states that he has had progressive*

*shortness of breath over the past several days. She states that today he was unable to wait any longer and needed to come to the hospital for further treatment. The patient's wife denies any fever, chills, productive cough or hemoptysis. The patient has had some mild nausea and vomiting about once per day.*

## Output

PATIENT NAME: Higher, Oberlander
PATIENT NUMBER: 00908983
INFO SOURCE: EMERGENCY ROOM NOTE
DATE: 9-3-91
AGE/GENDER: a 72 year old , white male
ILL_HEALTH/AFFLICTION: recurrent large cell lung carcinoma
DRUG/SUBSTANCE: CA
DRUG/SUBSTANCE: Tetracycline
BODY PART: the tube
SYMPTOM: any fever
SYMPTOM: productive cough or hemoptysis
SYMPTOM: mild nausea and vomiting

## Input

*Bouschor, Backman # 01029460 OPHTHALMOLOGY NOTE 8-28-91*
*This patient is a 15-year-old Latin American female who has had an erythematous, tender nodule in the medial third of the inferior lid of the right eye for approximately the last three days. The patient denies mattering of her eyes in the morning when she awakes. She states that this mass in her lower lid has induced some distortion of visual acuity. The patient has had a history of refractive error and has lost her glasses and needs new ones. No other visual complaints at this time.*

## Output

PATIENT NAME: Bouschor, Backman
PATIENT NUMBER: 01029460
INFO SOURCE: OPHTHALMOLOGY NOTE
DATE: 8-28-91
BODY PART: an erythematous , tender nodule
BODY PART: the inferior lid
BODY PART: her lower lid

## Input

*Nordell, Naples # 01633370 EMERGENCY ROOM NOTE 9-3-91*
*The patient is a 16-year-old male who was involved in an MVA. He was the driver and apparently fell asleep. He doesn't recall what happened. He has no complaints at all. He says that he doesn't hurt anywhere.*

## Output

PATIENT NAME: Nordell, Naples
PATIENT NUMBER: 01633370
INFO SOURCE: EMERGENCY ROOM NOTE
DATE: 9-3-91

AGE/GENDER: a 16-year-old male
ILL_HEALTH/AFFLICTION: MVA
ILL_HEALTH/AFFLICTION: hurt

## Input

*Boli,Bait # 02059026 EMERGENCY ROOM NOTE 9-2-91*
*CHIEF COMPLAINT: Temperature.*
*HPI: Bait is a 21 month old, white male who presents to the ER with his mother reporting that approximately six days ago the child was noted to have purulent nasal drainage and fever. At that time, he was evaluated by Pediatrics and begun on Amoxicillin. The clinic note notes infectious rhinitis and a temperature of 98.4. Symptomatic measures were prescribed for the diarrhea. A prescription for Amoxicillin was given for the nasal symptoms was prescribed by Dr. Riles. Since then the child has not been having any further diarrhea, but he has been having intermittent temperatures to 102-103, by report. The mother has been compliant giving the child Amoxicillin t.i.d. and Tylenol as needed. She has not been giving him any decongestant. The child has had normal activity, except that he has been fussy. He has been eating and drinking according to his mother. He has had no vomiting or abdominal pain.*

## Output

PATIENT NAME: Boli, Bait
PATIENT NUMBER: 02059026
INFO SOURCE: EMERGENCY ROOM NOTE
DATE: 9-2-91
AGE/GENDER: a 21 month old , white male
SYMPTOM: purulent nasal drainage and fever
DRUG/SUBSTANCE: Amoxicillin
ILL_HEALTH/AFFLICTION: infectious rhinitis
SYMPTOM: temperature of 98.4.
SYMPTOM: Symptomatic measures
SYMPTOM: the diarrhea
DRUG/SUBSTANCE: Amoxicillin
SYMPTOM: diarrhea
SYMPTOM: intermittent temperatures
DRUG/SUBSTANCE: Tylenol
SYMPTOM: vomiting
SYMPTOM: abdominal pain

*(Comment: Handling of negatives, as in "no vomiting or abdominal pain" or "doesn't hurt", has not been a focus and would still need attention.*

## Input

*Rassbach, Rhodefer # 02163335 PEDIATRIC NOTE 8/29/91*
*This 10 month old child comes in for a "right ear infection (?)". She was first seen on 8/12/91 after 3 weeks of runny nose, which was getting a little better by the time of being seen. There was a left middle ear effusion so Septa was prescribed. Things seem to be somewhat better, but then on 8/21/91 returned with fussiness and poor appetite. The left otitis was diagnosed and placed on Pediazole. Some stomachache resulted and so several days later Ceclor was started. Several days after this things seemed to be getting better. For the last four days the child has been fussy again and over the last two days has been runny nose, but*

*really no cough. The appetite is fair, drinking and voiding well. A little runny nose. Seems to be having some teething behavior. Stays with 4 other children at home that the mother keeps.*

## Output

PATIENT NAME: Rassbach, Rhodefer
PATIENT NUMBER: 02163335
INFO SOURCE: PEDIATRIC NOTE
DATE: 8-29-91
AGE/GENDER: 10 month old child
ILL_HEALTH/AFFLICTION: right ear infection
BODY PART: runny nose
BODY PART: Septa
SYMPTOM: fussiness
SYMPTOM: poor appetite
SYMPTOM: Some stomachache
BODY PART: runny nose
SYMPTOM: no cough
SYMPTOM: The appetite
BODY PART: A little runny nose

## Input

*Barfknecht, Nilsson # 02558166 EMERGENCY ROOM NOTE 9/4/91*
*CHIEF COMPLAINT:*
*Motor vehicle accident.*
*HISTORY OF THE PRESENT ILLNESS:*
*This 48 year old Hispanic male presents to the Emergency Department complaining of neck and left shoulder pain after being involved in a motor vehicle accident. The patient states that another car coming in the opposite direction pulled out in front of him and she struck the car. He was restrained with both shoulder and lap seat belts applied. No air bag was in the car. The patient denies striking the windshield, dashboard, or steering wheel. The patient has had no prior medical problems or injuries of the neck or shoulder, therefore, he does not have any neck pain, shortness of breath or other symptoms of cardiac problems, no peptic ulcer disease, no GI symptoms. No nausea, vomiting, or diarrhea.*

## Output

PATIENT NAME: Barfknecht, Nilsson
PATIENT NUMBER: 02558166
INFO SOURCE: EMERGENCY ROOM NOTE
DATE: 9-4-91
ILL_HEALTH/AFFLICTION: motor_vehicle_accident
SYMPTOM: neck and left shoulder pain
ILL_HEALTH/AFFLICTION: motor_vehicle_accident
BODY PART: the neck or shoulder
ILL_HEALTH/AFFLICTION: no peptic_ulcer disease
SYMPTOM: No nausea , vomiting
SYMPTOM: diarrhea

## Input

*Felise, Sr. Huhta # 02700175 EMERGENCY ROOM NOTE 9/2/91*
*The patient comes to the Emergency Room because of chest pain. He is 42 and this has been going on for about 2 weeks. He has a dozen or so episodes a day and frequently at night when he lays down to go to bed they are usually sharp and stabbing in the epigastrium lower sternal area and last for a few seconds. He does take antacids with relief. However, he states that he always has to take Alka Selzer or something like that after eating. The wife seems to think he was on Tagamet many years ago, but apparently has not had that problem recently. At his place of employment they have checked his cholesterol and it has been high at 275 to 300 range and he has been watching his diet trying to lower it. He is obese. His father had heart attack at age 60 and the patient basically leads a sedentary life. He is under a lot of stress at work.*

## Output

PATIENT NAME: Felise, Sr Huhta
PATIENT NUMBER: 02700175
INFO SOURCE: EMERGENCY ROOM NOTE
DATE: 9-2-91
SYMPTOM: chest_pain
BODY PART: lower sternal area
DRUG/SUBSTANCE: antacids
DRUG/SUBSTANCE: Alka Selzer
DRUG/SUBSTANCE: Tagamet
DRUG/SUBSTANCE: his cholesterol
ILL_HEALTH/AFFLICTION: heart_attack

## Input

*Windkam, Nilsson # 03586288 PEDI CLINIC NOTE 8-26-91*
*Eight and one-half month old brought in for a checkup. He has a history of hydroceles and hypospadias for which Dr. Palin has been following him. He is on Nutramigen as his formula and seems to be doing well on this.*

## Output

PATIENT NAME: Windkam, Nilsson
PATIENT NUMBER: 03586288
INFO SOURCE: PEDI CLINIC NOTE
DATE: 8-26-91
ILL_HEALTH/AFFLICTION: a checkup
ILL_HEALTH/AFFLICTION: hydroceles and hypospadias
DRUG/SUBSTANCE: Nutramigen

## Input

*Mcphee, Gerr # 03595084 EMERGENCY ROOM NOTE 9-3-91*
*The patient is a 31-year-old female who was involved in a trailer house fire last night. She comes in today because of headache. She is concerned about smoke inhalation. She says that she did feel like she inhaled quite a bit of smoke last night. She had no loss of consciousness, no nausea or vomiting. She did develop a mild headache several hours after the incident. She is a smoker.*

## Output

PATIENT NAME: Mcphee, Gerr
PATIENT NUMBER: 03595084
INFO SOURCE: EMERGENCY ROOM NOTE
DATE: 9-3-91
AGE/GENDER: a 31-year-old female
ILL_HEALTH/AFFLICTION: a trailer house fire
SYMPTOM: headache
DRUG/SUBSTANCE: smoke
SYMPTOM: no nausea or vomiting
SYMPTOM: a mild headache

## Input

*Nibyt, Dalla # 04376672 PEDIATRIC NOTE 8/29/91*
*This 19 month old child comes in with "diarrhea". The chart is not now available and the child is seen at the mother's request when she has brought her other child in. The child has been ill for about 3 days. The first signs of illness beginning day before yesterday with fever to 102.3 degrees and 10 small watery non bloody stools during the day. Had about 40 ounces of apple juice intake on that day. Had 3 episodes of vomiting on that day. Then yesterday he vomited twice, had 4 to 5 watery stools. Intake was about 24 ounces of Kool Aid and apple juice and he began to eat a little bit. Was less fussy yesterday. Today he has not really been fussy and is voiding adequately according to the mother but intake is uncertain as the child was at day-care before she was brought in. There has been a runny nose for the last several days, but no cough or spitting up. According to the mother a month or so ago the child weighed 22 to 23 pounds. Recent exposure is significant in that the child's sibling was hospitalized with diarrhea and dehydration about a week ago and also was later found to have Giardiasis, though stool cultures were negative. The mother in fact gave the child several doses of Thyroxin, 1/2 teaspoon yesterday and one dose of 1/2 teaspoon today.*

## Output

PATIENT NAME: Nibyt, Dalla
PATIENT NUMBER: 04376672
INFO SOURCE: PEDIATRIC NOTE
DATE: 8-29-91
AGE/GENDER: 19 month old child
SYMPTOM: diarrhea
ILL_HEALTH/AFFLICTION: illness
SYMPTOM: fever
SYMPTOM: 102.3 degrees
DRUG/SUBSTANCE: 10 small watery non bloody stools
SYMPTOM: vomiting
DRUG/SUBSTANCE: 5 watery stools
BODY PART: a runny nose
SYMPTOM: no cough
DRUG/SUBSTANCE: stool
BODY PART: Thyroxin

## Input

*Fenster, Terek # 04775358 EMERGENCY ROOM NOTE 9/2/91*
*Eleven years old and shot by 9 year old with a pellet gun. He comes with a bleeding wound in the groin. Actually it is in the inguinal area, just an inch or two to the base of the penis. Belly is soft and he was able to spontaneously void and the urine was dipstick negative. I put a marker over the entrance and had him take an x-ray but no foreign body was seen. I removed the marker and anesthetized the area up with 1/2 Marcaine with Epinephrine. He cleaned it thoroughly and there does not appear to be any foreign body present.*
*ASSESSMENT: Pellet wound. Left groin pain.*

## Output

PATIENT NAME: Fenster, Terek
PATIENT NUMBER: 04775358
INFO SOURCE: EMERGENCY ROOM NOTE
DATE: 9-2-91
ILL_HEALTH/AFFLICTION: a bleeding wound
BODY PART: the groin
BODY PART: the penis
BODY PART: Belly
DRUG/SUBSTANCE: the urine
BODY PART: no foreign body
DRUG/SUBSTANCE: 1 2 Marcaine with Epinephrine
ILL_HEALTH/AFFLICTION: Pellet wound
SYMPTOM: Left groin pain

## Input

Gwenn, Tenenbaum # 80942385 EMERGENCY DEPARTMENT NOTE 6-30-91
The patient is an 11-year-old, white male who hit his right index finger on a couch tonight. It has been tender and swollen since then. No other associated injuries. Right index finger has slight swelling and point tenderness at the PIP joint. The patient has pain with flexion but flexor function appears to be intact. Sensation is normal. X-rays of the right index finger were negative for fracture.

## Output

PATIENT NAME: Gwenn, Tenenbaum
PATIENT NUMBER: 80942385
INFO SOURCE: EMERGENCY DEPARTMENT NOTE
DATE: 6-30-91
AGE/GENDER: an 11-year-old , white male
BODY PART: his right index_finger
BODY PART: Right index_finger
SYMPTOM: swelling and point tenderness
BODY PART: the PIP joint
SYMPTOM: pain
BODY PART: the right index_finger
ILL_HEALTH/AFFLICTION: fracture

## Appendix B: Search Engine Example

### Medical Text Input

*CIM 12-3-91 S. C. Halsell,*
*OBSERVATION:*
*Bolland, Zisserman #17241245 CIM Clinic NOTE 12-3-91*

*Mrs. Bolland returns for a recheck. She has been to our ER several times since I last saw her. She apparently went Sunday, but those notes are not available. She tells me that they tried everything they knew to do and even treated her with IV Morphine to finally to break the headache. The headache got a little bit better, but she was wide awake for several days. She did not like this feeling. She is not a drug seeking type of person, but does have a somatization disorder. I would try and avoid narcotics in this patient in the future, if at all possible.*

*She tells me that her knee is doing a little bit better, although some days it seems to hurt a little bit more. She continues to take the 20 mg of Feldene per day.*

*She does have chronic recurrent migraine headaches in spite of her somatization disorder. She is currently on 50 mg of Tenormin per day and has been on that for about two weeks. She has noticed no significant improvement in her headaches. The Tenormin was used in conjunction with the recommendations made by Dr. Hideto.*

*She complains bitterly of recurrent neck discomfort. She points to the right posterior neck and she may have some osteoarthritis there and I think it is worthwhile getting some x-rays to make sure there is no obvious abnormalities.*

*On limited examination, blood pressure is 142/86 without orthostatic tilt.*

*ASSESSMENT:* *1. Somatization disorder.*
*2. Migraine headaches.*
*3. Neck discomfort.*
*PLAN: She is instructed to continue on her current medications and I have asked that she return in two weeks to see me. She will have a C-spine x-ray done now and if there are any significant abnormalities, I will contact her. She is dismissed to return in two weeks. I have encouraged her to contact me before trying to go to our ER.*

*S. C. Halsell, M.D./kdr*

---

*EMERGENCY DEPT 11-26-91 B. Dumes, D.O.,*
*OBSERVATION:*
*Bolland, Zisserman #17241245 EMERGENCY DEPT NOTE 11-26-91*

*The patient returns to ED after being seen in the clinic yesterday by Dr. Halsell. She has a headache she says, it is like the one she has had in the past and makes her neck sore and also her head hurts "all over especially up on top". It takes two medicines, but she does not know the name of either one.*

*EXAMINATION:*
*HEENT: Shows generalized redness of the upper respiratory mucosa. There is no tenderness to percussion or palpation over the maxillary or frontal sinuses.*
*NECK: Supple.*

*The remainder of the exam is unremarkable.*

*IMPRESSION: 1. Headache, uncertain etiology.*

*The patient also stated that she had been vomiting many times, thus did CBC and SMA-7. They were both unremarkable except CO2 was 21, glucose was 175, potassium was 5.4 with a slightly hemolyzed specimen.*

*DISPOSITION: The patient was given Nubain 5 mg. with Phenergan 25 mg. IV and felt much better. The pain was relieved. She was given Motrin 800 mg. to take one q. 6 h. after food p.r.m. pain, Phenergan 50 mg. suppositories one p.r. q. 6 h. p.r.m. severe nausea or vomiting or to enhance effect of pain medicine. Return if emergency. Discharge condition fair. The patient seemed happy with care and understood instructions.*

*B. Dumes, D.O./cjs*
*dd/dt: 11-26-91*
*Return for signature*

## Profiling the medical text

As described in last year's report (IRW95), (1) the entity selection is dependent upon profiling results along with a low breadth of meaning, (2) the subcategories provide a means of augmenting that original list with words of lower frequency that may otherwise have been eliminated, and (3) pattern matching is an active means by which unknown words occurring in known patterns (e.g., 35 milligrams of Tegretol) can be identified and assigned senses.

From the above medical text, the following list shows the full set of nouns immediately after profiling. The number following the noun is a count of the number of times the word occurred in the text. In all capital letters are shown the WordNet high level semantic categories in which each word falls. Occasionally the co-occurrence of one word with a second is also shown (indented on the next line).

```
abnormality     2  COGNITION ACT STATE ATTRIBUTE
assessment      1  COGNITION ATTRIBUTE POSSESSION
bit             3  QUANTITY EVENT FOOD OBJECT ARTIFACT
                little bit        3
bolland         3  REFERENCE PERSON
break           1  TIME ACT EVENT STATE
cim             2
clinic          2  ARTIFACT
condition       1  COGNITION COMMUNICATION STATE
conjunction     1  COMMUNICATION EVENT STATE ARTIFACT ATTRIBUTE RELATION
contact         2  PERSON COMMUNICATION ACT EVENT ARTIFACT RELATION
d.o.            1
                d.o./cjs    1
day             4  TIME
dd/dt           1
dept            2
                emergency dept  2
discharge       1  ACT SUBSTANCE EVENT PHENOMENON PROCESS
discomfort      2  FEELING STATE
                neck discomfort      2
disorder        3  STATE
                somatization disorder  3
disposition     1  COGNITION ACT ATTRIBUTE
dumes           2  REFERENCE PERSON
```

| | | |
|---|---|---|
| | b. dumes | 2 |
| ed | 1 | REFERENCE PERSON |
| effect | 1 | COGNITION COMMUNICATION ATTRIBUTE PHENOMENON |
| emergency | 3 | EVENT STATE ARTIFACT |
| er | 2 | LOCATION |
| etiology | 1 | COGNITION EVENT |
| examination | 2 | COGNITION COMMUNICATION ACT |
| feeling | 1 | COGNITION ACT STATE |
| feldene | 1 | SUBSTANCE |
| future | 1 | TIME COMMUNICATION |
| halsell | 3 | REFERENCE PERSON |
| head | 1 | LOCAT PERSON COGNIT COMMUNIC ACT BODY EVENT ARTIF SHAPE |
| headache | 7 | COGNITION STATE |
| heent | 1 | |
| hideto | 1 | REFERENCE |
| hurt | 2 | ACT EVENT FEELING STATE |
| impression | 1 | COGNITION COMMUNICATION ATTRIBUTE SHAPE |
| instruction | 1 | COMMUNICATION ACT |
| iv | 2 | QUANTITY |
| knee | 1 | BODY ANIMAL |
| m.d./kdr | 1 | |
| maxillary | 1 | |
| medication | 1 | ACT ARTIFACT |
| medicine | 2 | COGNITION ACT ARTIFACT |
| mg | 2 | QUANTITY SUBSTANCE |
| mg. | 4 | |
| migraine | 2 | STATE |
| | migraine headache | 2 |
| morphine | 1 | ARTIFACT |
| motrin | 1 | SUBSTANCE |
| mucosa | 1 | BODY |
| name | 1 | GROUP PERSON COMMUNICATION STATE |
| narcotic | 1 | ARTIFACT |
| nausea | 1 | FEELING STATE |
| neck | 5 | BODY FOOD OBJECT ARTIFACT |
| note | 3 | COMMUNICATION STATE POSSESSION |
| nubain | 1 | SUBSTANCE |
| observation | 2 | COGNITION COMMUNICATION ACT |
| orthostatic | 1 | |
| osteoarthritis | 1 | STATE |
| p.r. | 1 | |
| p.r.m. | 2 | |
| pain | 3 | COGNITION FEELING STATE |
| palpation | 1 | COGNITION ACT |
| patient | 5 | PERSON |
| percussion | 1 | ACT |
| person | 1 | COMMUNICATION |
| phenergan | 2 | SUBSTANCE |
| plan | 1 | COGNITION ARTIFACT |
| point | 1 | LOCAT TIME COGN COMMUN QUANT STATE OBJ ARTIF ATTRIB SHAPE |
| posterior | 1 | BODY |
| potassium | 1 | SUBSTANCE |
| recommendation | 1 | COMMUNICATION |
| recurrent | 2 | |
| redness | 1 | STATE ATTRIBUTE |
| remainder | 1 | ARTIFACT RELATION |
| return | 6 | COMMUNICATION ACT EVENT POSSESSION RELATION |
| right | 1 | GROUP ACT BODY ARTIFACT ATTRIBUTE RELATION |
| saw | 1 | COMMUNICATION ARTIFACT |

| | | |
|---|---|---|
| show | 1 | COMMUNICATION ACT EVENT ARTIFACT |
| signature | 1 | COMMUNICATION ATTRIBUTE |
| sinus | 1 | BODY |
| somatization | 3 | |
| | somatization disorder | 3 |
| sore | 1 | STATE |
| specimen | 1 | COGNITION BODY OBJECT |
| spite | 1 | FEELING |
| sunday | 1 | TIME |
| tenderness | 1 | FEELING STATE ATTRIBUTE |
| tenormin | 2 | SUBSTANCE |
| time | 2 | TIME EVENT ATTRIBUTE |
| type | 1 | GROUP COGNITION COMMUNICATION |
| upper | 1 | ARTIFACT |
| week | 3 | TIME |
| x-ray | 2 | ACT PHENOMENON |
| zisserman | 2 | REFERENCE PERSON |

Looking at these major categories, two problems are immediately apparent. First of all the categories are too broad. It is hard to see a strong connection between *tenderness* and *time* even though they are both *attributes*. The second problem is that many words have too many senses to provide meaningful information.

Looking at the original WordNet data for the words *headache* and *nausea*, we can see that there is a stronger connection than both being *states*. In fact, they are both *symptoms*. They also intersect in all of the categories superset to *symptom* (i.e., evidence, information, content, cognition), but for the moment, the level with most specificity and thus most knowledge will be of interest. Having determined that *symptom* is the subcategory of interest, the additional senses in *Cognition* and *Feeling* can be trimmed.

2 senses of headache

Sense 1 - STATE
headache, head ache, cephalalgia
=> pain, hurting
=> ***symptom***
=> evidence, grounds
=> information
=> content, cognitive content, mental object
=> cognition, knowledge
=> psychological feature

Sense 2 - COGNITION
concern, worry, headache, vexation
=> negative stimulus
=> stimulation, stimulus, stimulant, input
=> information
=> content, cognitive content, mental object
=> cognition, knowledge
=> psychological feature

---

2 senses of nausea

Sense 1 - STATE
nausea, sickness
=> ***symptom***
=> evidence, grounds
=> information
=> content, cognitive content, mental object
=> cognition, knowledge
=> psychological feature

Sense 2 - FEELING
nausea
=> disgust
=> dislike
=> feeling
=> psychological feature

Looking at each pair of words within the set of entities, the set of intersections can be examined. Printed here are those pairs whose intersection occurred at at least a depth of four. The depth threshold provides a means of ignoring intersections that may still be too general. In the above example, *content* at level 3 is probably still too general to be helpful. Unfortunately, the depth criteria is not entirely consistent across the net. At a depth of four for the first sense of *nausea* has *information* while the second sense has *disgust* which seems far more specific.

| | |
|---|---|
| assessment and observation | are kinds of *basic_cognitive_process* (2985532). (D=4) |
| assessment and pain | are kinds of *basic_cognitive_process* (2985532). (D=4) |
| assessment and palpation | are kinds of *basic_cognitive_process* (2985532). (D=4) |
| condition and effect | are kinds of *message* (3272071). (D=5) |
| condition and headache | are kinds of *information* (3026197). (D=4) |
| condition and instruction | are kinds of *message* (3272071). (D=5) |
| condition and migraine | are kinds of *information* (3026197). (D=4) |
| condition and name | are kinds of *message* (3272071). (D=4) |
| condition and nausea | are kinds of *information* (3026197). (D=4) |
| condition and observation | are kinds of *statement* (3325529). (D=6) |
| condition and pain | are kinds of *information* (3026197). (D=4) |
| condition and recommendation | are kinds of *message* (3272071). (D=5) |
| condition and redness | are kinds of *information* (3026197). (D=4) |
| condition and specimen | are kinds of *information* (3026197). (D=4) |
| day and sunday | are kinds of *calendar_day* (6627325). (D=4) |
| discomfort and hurt | are kinds of *suffering* (3624017). (D=4) |
| disposition and palpation | are kinds of *deed* (17087). (D=4) |
| disposition and percussion | are kinds of *deed* (17087). (D=4) |
| drug and medication | are kinds of *drug* (1611355). (D=4) |
| drug and medicine | are kinds of *drug* (1611355). (D=4) |
| drug and morphine | are kinds of *drug* (1611355). (D=4) |
| drug and narcotic | are kinds of *drug* (1611355). (D=4) |
| drug and upper | are kinds of *drug* (1611355). (D=4) |
| effect and future | are kinds of *idea* (3034185). (D=4) |
| effect and instruction | are kinds of *message* (3272071). (D=5) |
| effect and name | are kinds of *message* (3272071). (D=5) |
| effect and observation | are kinds of *message* (3272071). (D=5) |
| effect and person | are kinds of *idea* (3034185). (D=4) |
| effect and plan | are kinds of *idea* (3034185). (D=4) |
| effect and recommendation | are kinds of *message* (3272071). (D=5) |
| effect and type | are kinds of *idea* (3034185). (D=4) |

emergency and saw are kinds of *device* (1609642). (D=5)
er and mg are kinds of *chemical_element* (6414353). (D=4)
er and potassium are kinds of *chemical_element* (6414353). (D=4)
etiology and medicine are kinds of *discipline* (3084235). (D=5)
future and person are kinds of *grammatical_category* (3172257). (D=7)
future and plan are kinds of *idea*(3034185). (D=4)
future and type are kinds of *category* (3035986). (D=6)
headache and migraine are kinds of *headache* (6326486). (D=8)
headache and nausea are kinds of *symptom* (6321209). (D=6)
headache and pain are kinds of *pain* (6325495). (D=7)
headache and redness are kinds of *symptom* (6321209). (D=6)
headache and specimen are kinds of *information* (3026197). (D=4)
hurt and osteoarthritis are kinds of *ill_health* (6255096). (D=4)
hurt and percussion are kinds of *change_of_integrity* (131385). (D=4)
hurt and sore are kinds of *ill_health* (6255096). (D=4)
instruction and name are kinds of *message* (3272071). (D=5)
instruction and observation are kinds of *message* (3272071). (D=5)
instruction and plan are kinds of *creation* (1602007). (D=4)
instruction and recommendation are kinds of *message* (3272071). (D=5)
instruction and signature are kinds of *creation* (1602007). (D=4)
knee and neck are kinds of *body_part* (2756810). (D=5)
knee and posterior are kinds of *body_part* (2756810). (D=5)
knee and sinus are kinds of *body_part* (2756810). (D=5)
medication and medicine are kinds of *medicine* (1617942). (D=5)
medication and morphine are kinds of *drug* (1611355). (D=4)
medication and narcotic are kinds of *drug* (1611355). (D=4)
medication and upper are kinds of *drug* (1611355). (D=4)
medicine and morphine are kinds of *drug* (1611355). (D=4)
medicine and narcotic are kinds of *drug* (1611355). (D=4)
medicine and upper are kinds of *drug* (1611355). (D=4)
mg and potassium are kinds of *chemical_element* (6414353). (D=4)
migraine and nausea are kinds of *symptom* (6321209). (D=6)
migraine and pain are kinds of *pain* (6325495). (D=7)
migraine and redness are kinds of *symptom* (6321209). (D=6)
migraine and specimen are kinds of *information* (3026197). (D=4)
morphine and narcotic are kinds of *drug* (1611355). (D=4)
morphine and upper are kinds of *drug* (1611355). (D=4)
mucosa and specimen are kinds of *part* (2757486). (D=4)
name and observation are kinds of *message* (3272071). (D=5)
name and recommendation are kinds of *message* (3272071). (D=5)
name and signature are kinds of *name* (3182386). (D=5)
narcotic and upper are kinds of *drug* (1611355). (D=4)
nausea and pain are kinds of *symptom* (6321209). (D=6)
nausea and redness are kinds of *symptom* (6321209). (D=6)
nausea and specimen are kinds of *information* (3026197). (D=4)
neck and posterior are kinds of *body_part* (2756810). (D=5)
neck and sinus are kinds of *body_part* (2756810). (D=5)
observation and pain are kinds of *perception* (2987824). (D=5)
observation and palpation are kinds of *perception* (2987824). (D=5)
observation and recommendation are kinds of *message* (3272071). (D=5)
osteoarthritis and sore are kinds of *ill_health* (6255096). (D=4)
pain and palpation are kinds of *somesthesia* (2993451). (D=7)
pain and redness are kinds of *symptom* (6321209). (D=6)
pain and specimen are kinds of *information* (3026197). (D=4)

| | |
|---|---|
| palpation and percussion | are kinds of *touch* (48598). (D=5) |
| person and plan | are kinds of *idea* (3034185). (D=4) |
| person and type | are kinds of *category* (3035986). (D=6) |
| plan and signature | are kinds of *art* (2329483). (D=5) |
| plan and type | are kinds of *idea* (3034185). (D=4) |
| posterior and sinus | are kinds of *internal_structure* (2759523). (D=6) |
| redness and specimen | are kinds of *information* (3026197). (D=4) |
| remainder and upper | are kinds of *component* (1870155). (D=4) |

Tallying the categories of intersection results in the following list where C is the number of times that category was an intersecting category and D=# is the depth or specificity of this word in the net. The list is displayed in decreasing specificity D order. Proper names are being handled slightly differently because many are not in the semantic data set.

| Subcategory | Parent Categories in list if any | count/specificity |
|---|---|---|
| headache | pain#symptom#information | C=1 D=8 |
| grammatical_category | category#idea | C=1 D=7 |
| somesthesia | perception#basic_cognitive_process | C=1 D=7 |
| pain | somesthesia#symptom#perception#info | C=2 D=7 |
| symptom | information | C=7 D=6 |
| statement | message#art#information | C=1 D=6 |
| internal_structure | body_part | C=1 D=6 |
| category | idea | C=2 D=6 |
| message | | C=14 D=5 |
| body_part | | C=5 D=5 |
| device | deed | C=1 D=5 |
| art | | C=1 D=5 |
| touch | | C=1 D=5 |
| perception | idea#basic_cognitive_process | C=2 D=5 |
| name | message | C=1 D=5 |
| discipline | | C=1 D=5 |
| medicine | discipline#drug | C=1 D=5 |
| information | message | C=11 D=4 |
| drug | | C=14 D=4 |
| idea | | C=7 D=4 |
| component | | C=1 D=4 |
| part | | C=1 D=4 |
| creation | | C=2 D=4 |
| change_of_integrity | | C=1 D=4 |
| ill_health | | C=3 D=4 |
| chemical_element | | C=3 D=4 |
| deed | | C=2 D=4 |
| suffering | pain#symptom#information | C=1 D=4 |
| calendar_day | | C=1 D=4 |
| basic_cognitive_process | | C=3 D=4 |

Taking the preceding list and eliminating those that occur with a count=1 results in a much shorter list. For those occurring only once but whose parent category is also in the list, the parent category is incremented.

| Subcategory | Parent Categories in list if any | count/specificity |
|---|---|---|
| pain | somesthesia#symptom#perception#info | C=4 D=7 |
| symptom | information | C=7 D=6 |
| category | idea | C=2 D=6 |
| message | | C=16 D=5 |
| body_part | | C=6 D=5 |
| perception | idea#basic_cognitive_process | C=3 D=5 |
| information | message | C=11 D=4 |
| drug | | C=16 D=4 |
| idea | | C=7 D=4 |
| creation | | C=2 D=4 |
| ill_health | | C=3 D=4 |
| chemical_element | | C=3 D=4 |
| deed | | C=3 D=4 |
| basic_cognitive_process | | C=3 D=4 |

Running the original set of entities against this list and saving the subcategory in which each entity is a member results in a sorting by subcategory. It can additionally be thought of as a trimming of the word senses using the topic context.

At this point, one can sort out all the entities which are now identified and grouped as *drugs*, *information*, *body_part*, *etc.* As these groupings are obvious in the runtime *ask-it* example output, duplication is avoided here.

During the second pass with entities already identified, the sentence numbers in which each entity occurs are now saved. The starting and ending locations of each sentence are also saved for fast access.

## A Runtime Searching Engine

What follows is a session with a front end called "askit". It is mostly self-explanatory. The features to look for include the following:

(1) ability to show set of subcategories (option 'c'),
(2) ability to list entities which have a specified subcategory (e.g., 'l drug'),
(3) ability to show sentences containing specified entity or entities
(e.g. 'headache' brings up 5 sentences; 'headache' with 'morphine' brings up one sentence. (Use 's' to show current sentence set.)

>> *askit medical*
RUNNING: query -f medical.txt

----------index table---------

| | |
|---|---|
| assessment | [1] <COGNITION#basic_cognitive_process> |
| bolland | [3] <REFERENCE#female> |
| break | [1] <TIME#change_of_integrity> |
| cim | [2] <> |
| clinic | [2] <ARTIFACT> |
| condition | [1] <COGNITION#statement> |
| contact | [2] <PERSON#touch> |
| d.o. | [2] <> |
| day | [4] <TIME#calendar_day> |

| | |
|---|---|
| discharge | [1] <ACT#deed> |
| discomfort | [0] <FEELING#suffering> |
| disposition | [1] <COGNITION#deed> |
| dumes | [2] <REFERENCE> |
| effect | [1] <COGNITION#message> |
| emergency | [1] <EVENT#device> |
| emergency_dept | [2] <> |
| er | [2] <LOCATION#chemical_element> |
| etiology | [1] <COGNITION#discipline> |
| feeling | [1] <COGNITION#somesthesia> |
| feldene | [1] <SUBSTANCE#drug> |
| future | [1] <TIME#grammatical_category> |
| halsell | [3] <REFERENCE#doctor> |
| head | [1] <LOCATION#body_part> |
| headache | [5] <COGNITION#pain> |
| hideto | [1] <REFERENCE#doctor> |
| hurt | [2] <ACT#change_of_integrity> |
| impression | [1] <COGNITION#idea> |
| instruction | [1] <COMMUNICATION#message> |
| iv | [2] <QUANTITY> |
| knee | [1] <BODY#body_part> |
| medication | [1] <ACT#medicine> |
| medicine | [2] <COGNITION#discipline> |
| mg | [2] <QUANTITY#chemical_element> |
| mg. | [2] <> |
| migraine | [0] <STATE#headache> |
| migraine_headache | [2] <COGNITION#pain> |
| morphine | [1] <ARTIFACT#drug> |
| motrin | [1] <SUBSTANCE#drug> |
| mucosa | [1] <BODY#part> |
| name | [1] <GROUP#message> |
| narcotic | [1] <ARTIFACT#drug> |
| nausea | [1] <FEELING#symptom> |
| neck | [3] <BODY#body_part> |
| neck_discomfort | [2] <FEELING#suffering> |
| note | [3] <COMMUNICATION#statement> |
| nubain | [1] <SUBSTANCE#drug> |
| observation | [2] <COGNITION#statement> |
| osteoarthritis | [1] <STATE#ill_health> |
| p.r.m. | [1] <> |
| pain | [2] <COGNITION#somesthesia> |
| palpation | [1] <COGNITION#somesthesia> |
| patient | [5] <PERSON> |
| percussion | [1] <ACT#touch> |
| person | [1] <COMMUNICATION#grammatical_category> |
| phenergan | [2] <SUBSTANCE#drug> |
| plan | [1] <COGNITION#art> |
| point | [1] <LOCATION#message> |
| posterior | [1] <BODY#internal_structure> |
| potassium | [1] <SUBSTANCE#chemical_element> |
| recommendation | [1] <COMMUNICATION#message> |
| recurrent | [2] <> |
| redness | [1] <STATE#symptom> |
| remainder | [1] <ARTIFACT#component> |
| return | [6] <COMMUNICATION#idea> |
| right | [1] <GROUP#body_part> |
| saw | [1] <COMMUNICATION#device> |
| show | [1] <COMMUNICATION#message> |
| signature | [1] <COMMUNICATION#art> |

```
sinus                       [1] <BODY#internal_structure>
somatization_disorder       [3] <STATE#disorder>
sore                        [1] <STATE#ill_health>
specimen                    [1] <COGNITION#information>
sunday                      [1] <TIME#calendar_day>
tenderness                  [1] <FEELING#pain>
tenormin                    [2] <SUBSTANCE#drug>
time                        [2] <TIME#information>
type                        [1] <GROUP#category>
upper                       [1] <ARTIFACT#drug>
week                        [3] <TIME>
zisserman                   [2] <REFERENCE>
```

Subcategories:

person [Count=4]
drug [Count=8:1]
message [Count=6:5]
body_part [Count=4:2]
somesthesia :parent=perception#basic_cognitive_process# [Count=3:3]
pain :parent=somesthesia#symptom#information# [Count=3:3]
statement :parent=message#art#information# [Count=3]
chemical_element [Count=3]
grammatical_category :parent=category#idea# [Count=2]
symptom :parent=information# [Count=2:6]
internal_structure :parent=body_part# [Count=2]
device :parent=deed# [Count=2]
art [Count=2:5]
touch [Count=2]
discipline [Count=2:1]
information :parent=message# [Count=2:11]
idea [Count=2:3]
change_of_integrity [Count=2]
ill_health [Count=2]
deed [Count=2:2]
suffering :parent=pain#symptom#information# [Count=2]
calendar_day [Count=2]

```
Enter keyword (q=quit;n=new search;s=show;l=list;c=cats):l person
----------category table---------
contact                     [2] <PERSON#touch>
patient                     [5] <PERSON#person>
Enter keyword (q=quit;n=new search;s=show;l=list;c=cats):l reference
----------category table---------
bolland                     [3] <REFERENCE#female>
dumes                       [2] <REFERENCE>
halsell                     [3] <REFERENCE#doctor>
hideto                      [1] <REFERENCE#doctor>
zisserman                   [2] <REFERENCE>
Enter keyword (q=quit;n=new search;s=show;l=list;c=cats):l information
SUBCATS: pain
SUBCATS: statement
SUBCATS: symptom
SUBCATS: suffering
SUBCATS: headache
```

----------category table---------

| | |
|---|---|
| condition | [1] <COGNITION#statement> |
| discomfort | [0] <FEELING#suffering> |
| headache | [5] <COGNITION#pain> |
| migraine | [0] <STATE#headache> |
| migraine_headache | [2] <COGNITION#pain> |
| nausea | [1] <FEELING#symptom> |
| neck_discomfort | [2] <FEELING#suffering> |
| note | [3] <COMMUNICATION#statement> |
| observation | [2] <COGNITION#statement> |
| p.r.m. | [1] <> |
| recurrent | [2] <> |
| redness | [1] <STATE#symptom> |
| specimen | [1] <COGNITION#information> |
| tenderness | [1] <FEELING#pain> |
| time | [2] <TIME#information> |

Enter keyword (q=quit;n=new search;s=show;l=list;c=cats):l message
SUBCATS: statement
SUBCATS: information
SUBCATS: name
----------category table---------

| | |
|---|---|
| condition | [1] <COGNITION#statement> |
| effect | [1] <COGNITION#message> |
| instruction | [1] <COMMUNICATION#message> |
| name | [1] <GROUP#message> |
| note | [3] <COMMUNICATION#statement> |
| observation | [2] <COGNITION#statement> |
| point | [1] <LOCATION#message> |
| recommendation | [1] <COMMUNICATION#message> |
| show | [1] <COMMUNICATION#message> |
| specimen | [1] <COGNITION#information> |
| time | [2] <TIME#information> |

Enter keyword (q=quit;n=new search;s=show;l=list;c=cats):l body_part
SUBCATS: internal_structure
----------category table---------

| | |
|---|---|
| head | [1] <LOCATION#body_part> |
| knee | [1] <BODY#body_part> |
| neck | [3] <BODY#body_part> |
| posterior | [1] <BODY#internal_structure> |
| right | [1] <GROUP#body_part> |
| sinus | [1] <BODY#internal_structure> |

Enter keyword (q=quit;n=new search;s=show;l=list;c=cats):l idea
SUBCATS: grammatical_category
SUBCATS: category
SUBCATS: perception
----------category table---------

| | |
|---|---|
| future | [1] <TIME#grammatical_category> |
| impression | [1] <COGNITION#idea> |
| person | [1] <COMMUNICATION#grammatical_category> |
| return | [6] <COMMUNICATION#idea> |
| type | [1] <GROUP#category> |

```
Enter keyword (q=quit;n=new search;s=show;l=list;c=cats):l chemical_el
----------category table---------
er                              [2] <LOCATION#chemical_element>
mg                              [2] <QUANTITY#chemical_element>
potassium                       [1] <SUBSTANCE#chemical_element>

Enter keyword (q=quit;n=new search;s=show;l=list;c=cats):l drug
SUBCATS: medicine
----------category table---------
feldene                         [1] <SUBSTANCE#drug>
medication                      [1] <ACT#medicine>
medicine                        [2] <COGNITION#drug>
morphine                        [1] <ARTIFACT#drug>
motrin                          [1] <SUBSTANCE#drug>
narcotic                        [1] <ARTIFACT#drug>
nubain                          [1] <SUBSTANCE#drug>
phenergan                       [2] <SUBSTANCE#drug>
tenormin                        [2] <SUBSTANCE#drug>
upper                           [1] <ARTIFACT#drug>
Enter keyword (q=quit;n=new search;s=show;l=list;c=cats):phenergan
phenergan; SUBSTANCE; #drug; ; ; ;
---------- table pairs---------
                effect                  66:
                iv                      64:
                medicine                66:
                mg.                     64:             66:
                motrin                  66:
                nausea                  66:
                nubain                  64:
                p.r.m.                  66:
                pain                    66:
                patient                 64:
                phenergan               64:             66:
Enter keyword (q=quit;n=new search;s=show;l=list;c=cats):s
                64:
The patient was given Nubain 5 mg. with
Phenergan 25 mg. IV and felt much better.
                66:
She was given Motrin 800 mg. to take one q. 6 h. after food p.r.m.
pain, Phenergan 50 mg. suppositories one p.r. q. 6 h. p.r.m. severe
nausea or vomiting or to enhance effect of pain medicine.

Enter keyword (q=quit;n=new search;s=show;l=list;c=cats):n
Enter keyword (q=quit;n=new search;s=show;l=list;c=cats):tenormin
tenormin; SUBSTANCE; #drug; used; ; ;
---------- table pairs---------
                day                     17:
                hideto                  19:
                mg                      17:
                recommendation          19:
                tenormin                17:             19:
                week                    17:
Enter keyword (q=quit;n=new search;s=show;l=list;c=cats):s
```

17:
She is currently on 50 mg of Tenormin per
day and has been on that for about two weeks.
19:
The Tenormin was used in
conjunction with the recommendations made by Dr. Hideto

Enter keyword (q=quit;n=new search;s=show;l=list;c=cats):n
Enter keyword (q=quit;n=new search;s=show;l=list;c=cats):headache
headache; COGNITION; #information; ; ; ;
Also consider migraine_headache?y
---------- table pairs---------

| | | | | | |
|---|---|---|---|---|---|
| break | 7: | | | | |
| day | 8: | | | | |
| etiology | 58: | | | | |
| head | 45: | | | | |
| headache | 7: | 8: | 18: | 45: | 58: |
| hurt | 45: | | | | |
| iv | | 7: | | | |
| medicine | | 45: | | | |
| migraine_headache | | 16: | 28: | | |
| morphine | | 7: | | | |
| name | | 45: | | | |
| neck | | 45: | | | |
| recurrent | | 16: | | | |
| sore | | 45: | | | |

Enter keyword (q=quit;n=new search;s=show;l=list;c=cats):s
7:
She tells me that they tried everything
they knew to do and even treated her with IV Morphine to finally
to break the headache.
8:
The headache got a little bit better, but she
was wide awake for several days.
16:
She does have chronic recurrent migraine headaches in spite of her
somatization disorder.
18:
She has noticed no
significant improvement in her headaches.
28:
2. Migraine headaches. 3
45:
She has a headache she says, it is like the one she
has had in the past and makes her neck sore and also her head hurts
"all over especially up on top". It takes two medicines, but she
does not know the name of either one.
58:
1. Headache, uncertain etiology

Enter keyword (q=quit;n=new search;s=show;l=list;c=cats):n
Enter keyword (q=quit;n=new search;s=show;l=list;c=cats):headache
headache; COGNITION; #information; ; ; ;
Also consider migraine_headache?y

---------- table pairs---------

| | | | | | |
|---|---|---|---|---|---|
| break | | 7: | | | |
| day | | 8: | | | |
| etiology | | 58: | | | |
| head | | 45: | | | |
| headache | 7: | 8: | 18: | 45: | 58: |
| hurt | | 45: | | | |
| iv | | 7: | | | |
| medicine | | 45: | | | |
| migraine_headache | | 16: | 28: | | |
| morphine | | 7: | | | |
| name | | 45: | | | |
| neck | | 45: | | | |
| recurrent | | 16: | | | |
| sore | | | 45: | | |

Enter keyword (q=quit;n=new search;s=show;l=list;c=cats):morphine
morphine; ARTIFACT; #drug; IV Morphine; ; ;
---------- table pairs---------

| | |
|---|---|
| break | 7: |
| headache | 7: |
| iv | 7: |
| morphine | 7: |

Enter keyword (q=quit;n=new search;s=show;l=list;c=cats):s
7:
She tells me that they tried everything
they knew to do and even treated her with IV Morphine to finally
to break the headache.

Enter keyword (q=quit;n=new search;s=show;l=list;c=cats):q

**end of session**

DISTRIBUTION:

| | | | |
|---|---|---|---|
| 2 | Kim S. Culp, M.D.<br>Scott & White Hospital<br>2401 South 31 Street<br>Temple, TX 76508 | | |
| 1 | MS | 0320 | Chuck Meyers, 1011 |
| 1 | | 0459 | John Taylor, 5336 |
| 1 | | 0473 | William Cook, 5312 |
| 1 | | 0449 | Susan Copus., 9411 |
| 3 | | 0473 | Sharon Deland, 5312 |
| 1 | | 1138 | Sharon Chapa, 9411 |
| 6 | | 0449 | Nancy Irwin, 9411 |
| 1 | | 0829 | Stephen Crowder, 12323 |
| 1 | | 0573 | Huri Fraley, 5913 |
| 1 | | 0573 | Kathie Hiebert-Dodd, 5913 |
| 1 | | 0575 | Joyce Van Berkel, 5951 |
| 1 | | 0570 | Nancy Pruett, 5951 |
| 1 | | 0449 | Judy Moore, 9415 |
| 1 | | 0761 | Larry Desonier, 5822 |
| 1 | | 0431 | Sam Varnado, 9400 |
| 1 | | 0451 | Brian Cloer, 9403 |
| 1 | | 9018 | Central Technical Files, 8523-2 |
| 5 | | 0899 | Technical Library, 4414 |
| 2 | | 0619 | Review and Approval Desk, 12630<br>For DOE/OSTI |