THE SYSTEM FOR NEUROLOGICAL ANALYSIS OF PATIENT SYMPTOMS: AN INTEGRATED ARTIFICIAL INTELLIGENCE PROTOTYPE

Jeffrey L. Sponsler MD[1], Gayane Goltukhchyan[2], Frances Van Scoy PhD[2], James Culberson PhD[3]
1 Case Western Reserve University Department of Neurology, Cleveland, Ohio, USA 44106 www.jeffsponsler.com;
2 West Virginia University Department of Computer Science and Engineering;
3 West Virginia University Department of Anatomy, Morgantown WV, USA 26506

Abstract

Objectives: To create a stroke expert system prototype that reads a patient text file, computes a disease location, and presents brain graphics showing the lesion. *Methods*: Coded in Common Lisp, The Natural Language Module reads a text file, parses, and produces parse trees. An Information Extraction Module employs rules to extract data from parse trees and to store these into a database. A Neuro-Anatomic Atlas has been created to represent clinically relevant nervous system structures. A Localization Module employs rules and methods to map from a clinical finding to an anatomic structure. The Graphics Module presents representations of brain structures corresponding to the localization. A 3D Brain Model has been created by scanning a specimen and is a part of this module. *Results*: Text files representing stroke syndromes are analyzed by the system which produces localization hypotheses and graphical representations of the structures. Formal controlled trials of program behavior are planned. *Conclusions*: The prototype demonstrates feasibility and should, as new knowledge is incorporated, evolve into a utility for diagnosis, instruction, and experimentation. Formal testing will be required in later stages of the project.

Keywords

Expert system, neurology, stroke, neuroanatomy, natural language processing

1. Introduction

The early MYCIN system automated bacterial species identification [1]. The PROSPECTOR program used hypothesis analysis to discover mineral deposits [2]. The systems DXPLAIN [3] and QMR [4] contain knowledge of internal medicine.

A neurology expert system should emulate the four stages that a clinician uses. The first stage is information gathering (history, physical examination, laboratory values, and interpretation of imaging studies) stored in a report. The term *finding* is defined as a unit of information from this discovery process. The second stage, localization, requires knowledge of anatomy and produces nervous system structures that may be injured and thus may account for the findings accumulated in the first step. The third stage, diagnosis, requires inputs from the prior two stages and creates a list of diseases that may be responsible for the structural injury and the patient symptoms. The fourth step, work-up, includes further testing to support or disprove a diagnostic possibility.

The *System for Neurological Analysis of Patient Symptoms* (SYNAPS) prototype has been completed and this report documents the progress made and future development. The goals of SYNAPS include 1) diagnose neurologic disease, 2) represent most nervous system structures in an organized knowledge base, 3) provide a workbench for development of instructional and investigational interfaces, and 4) integrate graphical models of nervous system structures with the anatomic knowledge base. The diagnostic emphasis in this prototype is **stroke** (brain infarct secondary to occlusion of a cerebral artery).

1.1. Heuristics

We propose heuristics, the Common Arterial Supply Heuristic, the Common Parent Structure Heuristic, and the Finding to Subpathway Heuristic, to describe the basis for anatomic localization in SYNAPS. These are described below.

2. The SYNAPS Modules

The components of SYNAPS include a report parser, a database population module, a localization module, and a graphics presentation module (Figures 1 and 2).

Findings Menu Interface. A menu based findings interface is provided as another input mechanism. This interface has been coded in Lisp and in a client/server browser system.

2.1. HPARSER: The Natural Language Processing Module

The architecture of HPARSER [5] consists of a Parsing Module and a Database Constraint Processor. Knowledge inputs include the *Unified Medical Language System* (UMLS) Lexicon [6], *English Grammar Rules*, *Medical Grammar Rules*, and *Database Constraint Rules*. HPARSER is a recursive transition network system [7]. An input sentence is parsed by the Parsing Module (using the Lexicon and Grammar Rules). The medical grammar rules encode a set of commonly used phrases. The Database Constraint Processor, using the Database Constraint Rules, extracts information from each parse tree and stores such information into a database table. Figure 1 illustrates the modules, the knowledge bases, and the data path of HPARSER.

2.2. NeuroAnatomic Atlas: Module storing nervous system structures.

The NeuroAnatomic Atlas [8] consists of objects and relationships. Important definitions for this report include the following: A *nucleus* is a discrete collection of neurons, a *connection* is a bundle of axons connecting two nuclei, and a *subpathway* is a linear set of connections. A *dermatome* is an area of skin innervated by a specific spinal nerve. *Vibration* is a sensation involving perception of physical oscillation on the skin surface. The fifth cervical spinal cord nerve root is called *C5*. Currently the NAA has 2178 entries and includes these classes: brainstem nuclei, white matter tracts, muscles, the brachial plexus, dorsal root ganglia, the spinal cord, cortical areas, the basal ganglia, and arteries. Also represented are sensory pathways (the *dorsal column medial lemniscus* system) and motor pathways (the *corticospinal tract*). Structures are assigned a specific function in the atlas. For example, the function of the lateral rectus muscle is to abduct the eye. NAA is encoded using the Common Lisp Object System (CLOS) [9].

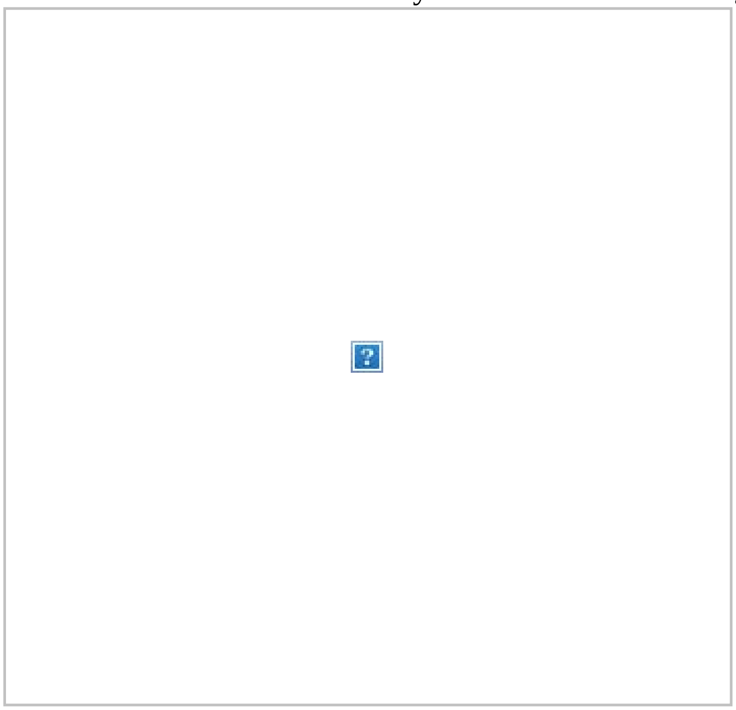

Figure 1. HPARSER Architecture. Sentences are processed resulting in parse trees that are further analyzed yielding data storage into a database.

2.3. Localization Algorithms.

The Localization Algorithms are a set of methods that process information from the patient database to produce a list of localizations. The most important localization step is mapping from finding to anatomic structure. Several heuristics support additional analysis of these findings and structures.

<u>The Common Arterial Supply Heuristic</u>: If anatomic structures share the same arterial supply, then malfunction of that artery may be a contributing factor in the disease process.

<u>The Common Parent Structure Heuristic</u>: If two structures S1 and S2 exist within a larger common parent structure Sp, and S1 and S2 exhibit dysfunction, then Sp is a possible site of injury. For example, the C5 vibration sense subpathway and C5 motor subpathway share a common region in the cerebral cortex near the central sulcus. This region may be considered as a localization hypothesis. Another example of Common Parent Structure Heuristic is the *lateral medullary syndrome* (constricted pupil, ipsilateral facial sensory dysfunction, contralateral loss of pain sensation, dysequilibrium) that is secondary to stroke in the lateral medulla. The trigeminal and the spinothalamic systems are adjacent in this structure.

<u>The Finding To Subpathway Heuristic</u>. If function of a subpathway is impaired (i.e., a finding), then disease process has caused injury to one or more components of this subpathway.

3. Description of Data Flow

The flow of data from the free text patient report file to anatomic localization is demonstrated via a brief example. This process requires knowledge bases (UMLS Lexicon, Grammar Rulebase, Medical Grammar Rulebase, Database Constraints, Neuroanatomic Atlas) and processing modules to apply the knowledge to each step in data manipulation. To illustrate, a simplistic file is used here.

3.1. Patient Report File

The patient file is an ASCII text file containing a typical patient report. The example sentences are:

```
Right biceps strength is 2/5.
Vibration in the right C5 dermatome is absent.
```

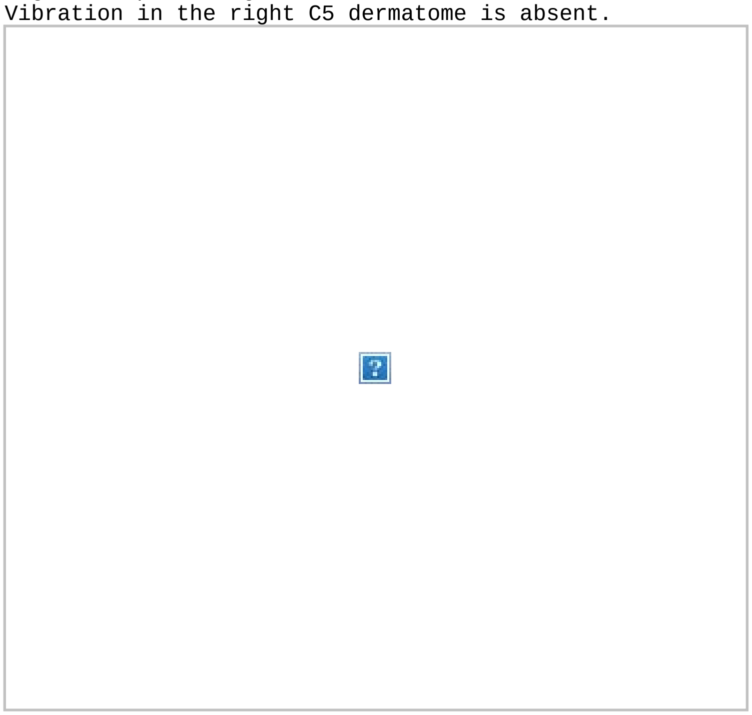

Figure 2. Modules and data flow of the localization and diagnostic components of the Localization Module.

3.2. Creating a Set Of Word Tokens

The initial step involves reading the text file and identifying word elements, punctuation, numerals, and sentence delimiters. The result is a set of sentence lists. Each list contains a set of objects of class *token*. In this report, objects are represented textually by the convention of **<*object*>**.

```
(<right> <biceps> <strength> <is> <|numeral|> <|slash|> <|numeral|>)
(<vibration> <in> <the> <right> <C5> <dermatome> <is> <absent>)
```

3.3. Generating Parse Trees

The Parser Module processes each the sentence token list and produce a parse tree. The software uses a rule-based depth-first parsing algorithm.

Software has been developed (based on Common Graphics utilities) to present graphical representation of the parse trees for the user.

3.4. Populating The Patient Database

The next phase involves data extraction from the parse trees. The Database Constraint Rules that associate grammar rules and database tables are applied and findings are stored in the database. A class *location-finding* is instantiated when information is extracted from a parse tree. This information is placed into slots *side*, *location*, and *value*. At the conclusion of this step, the database tables will be populated with findings.

```
Analyzing sentence 84: (<right> <biceps> <strength> <is> <|numeral|> <|slash|> <|numeral|>)
Processing Unilateral-Search-Constraint RIGHT-BICEPS-STRENGTH
Storing <Finding GRADE-TWO> into NEURO-EXAM/RIGHT-BICEPS-STRENGTH

Analyzing sentence 93: (<vibration> <in> <the> <right> <C5> <dermatome> <is> <absent>)
Processing Multiple-Search-Constraint RIGHT-C5-VIBRATION Storing <Finding ABSENT> into NEURO-EXAM/RIGHT-C5-VIBRATION
```

3.5. The Findings Table

To display findings to the user of SYNAPS, an interactive Findings Table has been developed. The table contains information about clinical findings and associated subpathways. Figure 3 describes the contents of one example row in this table.

3.6. The Connections Table

To display subpathway connections to the user of SYNAPS, a Connections Table window has been developed. Mouse selection in the Findings Table of a subpathway object produces a new window, the Connections Table. Figure 4 describes the contents.

Column	Description	Example
Sentence	The sentence (from the patient report) from which the finding information is derived.	<Sentence 97 (<vibration> <sensation> <in> <the> <right> <c5> <dermatome> <is> <absent>)>
Database Constraint	The database constraint that was used to identify a parse tree pattern and to extract the finding from the tree.	<DB-Constraint RIGHT-C5-VIBRATION>
Location	Anatomic structure that is identified in the sentence.	Dermatome-C5
Side	Left, right, or bilateral	Right
Modality	Function of the anatomic structure.	Vibration
Value	Description of the function of the anatomic structure based on the sentence content.	Absent
Subpathway	The subpathway object that is associated with the finding identified in the sentence. This is derived from 1) the atlas where the structure is assigned a function and 2) the data in columns Location, Side, and Modality.	<SENSORY SUBPATHWAY c5-vibration-right>

Figure 3. The Findings Table is described here.

3.7. The Localization Step

The final step in the data path involves iterating through all findings in the database tables. The location-finding class stores side (left or right), locus (C5 dermatome), modality (e.g., vibration), and value (e.g., normal or absent).

The atlas stores knowledge relating specific functions (vibration) to a subpathway (vibration subpathway). The functions serve as keys into the NAA. Such keys provide a mapping from patient findings (abnormalities of function) to subpathways in the nervous system.

Example: The sentence “Vibration sense in right C5 dermatome is absent” maps

(via the Finding To Subpathway Heuristic) to the C5 Cortical Sensory Area. The sentence “Right biceps strength is 4/5” maps to the cortical motor area. These areas are contiguous on the cortex. By the Common Parent Structure Heuristic, localization is the region of the cerebral cortex near the left central sulcus (specifically the sensory and motor areas). By the Common Arterial Supply Heuristic, occlusion of the left middle cerebral artery is a consideration.

3.8. Anatomic Image Library

The prototype at present contains a set of 2-D and 3-D image files that are catalogued in a table. The images are indexed by marked anatomic areas (see Figure 5). The interface presents the list of localizations. Selecting a hypothesis displays the image.

4. Discussion

Our prototype demonstrates feasibility of the SYNAPS approach to a neurology expert system. The fusion of several technologies is a key feature of this system. Standard AI techniques (natural language processing, object-oriented programming, rule-based diagnosis) have been applied.

Future Work. Each knowledge base in this system will require continued evolution. Adding new structures to the atlas is planned so that all clinically relevant objects are included. The parsing module will require further work primarily in expanding the rule bases. The graphics module will be enhanced so that 2D magnetic resonance images and 2D Visual

Human [10] images are employed to show internal structures.

The diagnostic module is in evolution. A Lisp module, Prolisp, is based on Prolog [11] and SRI's New Automated Reasoning Kit [12]. Prolisp has been employed to localize temporal lobe epilepsy. Formal trials of SYNAPS comparing human expert opinion and SYNAPS behavior are planned.

Column	Description	Example
Connection-ID	Unique symbol for object	Connection-1
Source	Source (set of neurons) of connection's axons	<SENSORY-NUCLEUS vpl c5 vibration left>
Destination	Postsynaptic destination of connection's axons	<AREA sensory area 1 c5 left>
Tract	Fiber tract through which axons of connection traverse	<CAPSULE posterior limb internal capsule left>
Artery	Artery that serves the tract	<ARTERY lateral striate artery left>

Figure 4. The Connections Table is described here.

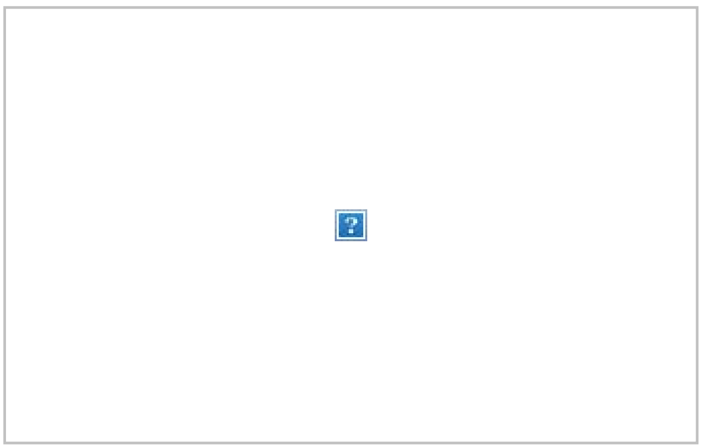

Figure 5. In this image (from a scanned specimen), the left motor cortex (precentral gyrus) for the arm is marked.

5. References

[1] Shortliffe, E.H.. *MYCIN: Computer-Based Medical Consultations*. New York: American Elsevier, 1976.

[2] Duda, Hart, and Nilsson, 1976. Subjective Bayesian Methods for Rule-Based Inference Systems, in *Proceedings of the 1976 National Computer Conference*, AFIPS Press (June 1976).

[3] Sumner II, W. A review of Iliad and Quick Medical Reference for primary care providers. Two diagnostic computer programs. *Arch Fam Med* 1993 Jan;2:87-95.

[4] Miller R, Masarie FE, Myers JD. *Quick Medical Reference (QMR) for diagnostic assistance*. MD Comput 1986 Sep-Oct;3(5):34-48.

[5] Sponsler, Jeffrey L. 2001. HPARSER: Extracting formal patient data from free text history and physical reports using natural language processing software. *Proceedings of the American Medical Informatics Association Annual Symposium*, Washington DC, Nov 3-11, 2001.

[6] McCray, A., S. Srinivasan, A. Browne, 1994. Lexical methods for managing variation in biomedical terminologies. *Proc of the 18th Annual Symposium on Computer Applications in Medical Care*, pp. 235-239.

[7] Gazdar, G, C. Mellish, 1989. *Natural Language Processing in LISP* (Addison-Wesley, 1989).

[8] Sponsler, Jeffrey, Frances Van Scoy, Doru Pacurari, 2002. An object oriented neuroanatomic atlas, *Proc of Conference of Artificial Intelligence and Soft Computing*, Banf Canada, July 2002.

[9] Keene, Sonya, 1989. Object-Oriented Programming in Common Lisp: A Programmer's Guide to CLOS, Addison-Wesley, Reading, PA.

[10] Toh, Young, Robert B. Falk, James S. Main, 1996. Digital Brain Atlas Using the Visible Human Project Data: The

Visible Human Project Conference, *Proc October 7 & 8, 1996 National Institutes of Health*, Natcher Conference Center, Bethesda, Maryland USA.

[11] Clocksin, WF, CS Mellish, 1987. *Programming in Prolog*. Spring-Verlag, Berlin, chapter 10.

[12] Stickel, Mark, 2003. URL: http://www.ai.sri.com/~stickel/snark.html.