

Все программные модули работают под управлением ОС Windows 95/98/Me/2000/XP.

И. А. Коломойцева

ФОРМИРОВАНИЕ БАЗЫ ЗНАНИЙ И ПОСТАНОВКА ДИАГНОЗА НА ОСНОВЕ ЕСТЕСТВЕННО-ЯЗЫКОВЫХ ПСИХОЛОГИЧЕСКИХ ТЕКСТОВ

В любой экспертной системе можно выделить четыре основных функциональных блока:

- базу знаний;
- блок вывода заключений;
- блок "объяснения" вывода заключений;
- блок наполнения и редактирования базы знаний.

База знаний является важной частью экспертной системы, качество наполнения которой серьезно влияет на всю работу системы. Практические методы получения знаний можно разделить на два больших вида: коммуникативные и текстологические. Основной принцип деления связан с источником знаний. В первом случае таким источником является эксперт, а во втором – литература, документы, учебники [1]. Именно текстологический способ извлечения знаний может быть использован для создания автоматизированной системы построения базы знаний.

Процесс автоматизации принятия решения в экспертных системах невозможен без привлечения информации, которая не может быть выражена количественно. Это семантическая (смысловая) информация. Такую информацию возможно извлечь из естественно-языковых текстов, в случае медицинских экспертных систем знания извлекаются из специальных медицинских текстов, зафиксированных рассказов врачей-экспертов о различных проявлениях заболеваний, методов их лечения и т. п. [2, 3].

На протяжении последних двух-трех десятилетий многие исследователи, занимающиеся проблемами автоматической обработки текста и «понимания» естественного языка, получили ряд интересных результатов [4, 5, 6].

Анализ текста предполагает наличие нескольких фаз:

- морфологический анализ;
- синтаксический анализ;
- семантический анализ;
- прагматический анализ.

Фаза семантического анализа текста, представленная в этой статье, использует методику, описанную в [5, 6], основанную на выделении семантических связей в естественно-языковом тексте.

В табл. 1 приведены семантические связи, наиболее часто встречающиеся в медицинских текстах.

Таблица 1

Наименование семантической связи	Обозначение семантической связи	Описание семантической связи
Генеративная	Gen	Один компонент обозначает лицо или предмет, принадлежащий некоторой совокупности, категории, обозначаемой вторым компонентом
Инструментальная	Ins	Один компонент обозначает орудие действия, обозначаемого другим компонентом
Директивная	Dir	Один компонент обозначает путь, направление второго компонента
Каузальная	Cous	Один компонент обозначает причину появления другого компонента спустя какое-то время
Комитативная	Com	Один компонент обозначает сопровождающее другой компонент действие, сопутствующий предмет, сопровождающее лицо
Коррелятивная	Cog	Один компонент выражает возможность наблюдения другого компонента или соответствия предмета другому предмету, назначению

В табл. 2 представлены примеры предикторов, встречаемых в естественно-языковом медицинском тексте, и указаны семантические связи, к которым они относятся. В результате анализа предложений, взятых из 20 описаний психических заболеваний, было установлено, что 28% из них содержат предикторы, относящиеся к каузальной связи, 25% – к комитативной, 17% – к коррелятивной и к результативной и 4 – к генеративной, директивной и инструментальной.

Таблица 2

Признаки болезни	Связь
начинается в возрасте	генеративная
	комитативная
обнаруживается	комитативная
проявляется	комитативная
сопровождаются симптомами	комитативная
происходит	комитативная
протекает	коррелятивная
наблюдаются	коррелятивная
имеют черты	коррелятивная
характерная особенность	коррелятивная
свойственно	коррелятивная
больные становятся	казуальная
сопровождаются	казуальная
возникают	казуальная

Создаваемая база знаний представляет собой сеть фреймов. Структура основного фрейма и примеры значений слотов представлены в табл. 3.

Таблица 3

Основные сведения о заболевании					
Название заболевания	Признаки	Сопутствующие заболевания	Последствия	Причины	Методы лечения
Церебральный тромбангиит	цереб.	нарушение памяти, вялость, беспомощность			
Аффективный психоз	аффек.	усиление тревоги при перемене места	шизофрения		амбулаторно
Миоклонус-эпилепсия	миок.	капризность, плаксивость, назойливость	слабоумие, сумеречное состояние сознания	наследственность	

В качестве значения слота «Признаки» используются имена фреймов, содержащих основные признаки заболевания. Примеры таких фреймов приведены в табл. 4, 5, 6.

Таблица 4

аффек						
Признак 1	Признак 2	Признак 3	Признак 4	Признак 5	Признак 6	Признак 7
повышенное настроение	двигательное возбуждение	самообвинение	суточные колебания настроения	расстройство сна	запоры	бредовые идеи

Таблица 5

аффек						
Признак 1	Признак 2	Признак 3	Признак 4	Признак 5	Признак 6	Признак 7
повышенное настроение	двигательное возбуждение	самообвинение	суточные колебания настроения	расстройство сна	запоры	бредовые идеи

Таблица 6

миок					
Признак 1	Признак 2	Признак 3	Признак 4	Признак 5	Признак 6
эпилептические припадки	возраст 10-16	неврологические нарушения	дистрофия	нарушение памяти	ослабление внимания

При диагностировании проверяемому предлагается в произвольной форме рассказать о своем самочувствии. По его рассказу строится фрейм его психического состояния. Посредником при построении данного фрейма является переводчик, который преобразует слова, сказанные человеком, в термины, понятные базе зна-

ний. Пример такого перевода термина «прожорливость» на естественный язык приведен в табл. 7.

Далее фрейм пациента, проверяемого на принадлежность к группе риска, сравнивается с фреймами заболеваний, имеющихся в базе знаний. Если в результате проверки обнаруживается совпадение с каким-то заболеванием по прямым признакам на $\geq 80\%$ и по сопутствующим на $\geq 60\%$, то считается, что пациент принадлежит к группе риска по этому заболеванию и требует более детального обследования.

Таблица 7

Прожорливость
постоянно ем
всегда ем
всегда хочу есть
ем все подряд и всегда
всегда жую что-нибудь
стал(а) много есть
ем все время

В качестве примера рассмотрим следующее описание состояния пациента: «Хочу постоянно спать, все забываю, часто нервничаю, иногда болит и кружится голова, часто тошнит, ничего не получается, мне 26 лет». Пропустив данный набор фраз через переводчик получаем следующие признаки: сонливость, нарушение памяти, головная боль, головокружение, рвота, возраст 26 лет, беспомощность, невроз. Проверка по прямым и сопутствующим признакам дала диагноз: «церебральный тромбангиит».

Система, представленная в данной статье, существенно упрощает обработку естественно-языкового медицинского текста и позволяет ее автоматизировать. С другой стороны, с помощью данной системы предварительный диагноз может поставить и не очень высоко квалифицированный в области психиатрии врач.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Базы знаний интеллектуальных систем / Гаврилова Т.А., Хорошевский В.Ф. – СПб: Питер, 2000. – 384 с.
2. Коломойцева И.А. Математическая модель представления естественно-языкового медицинского текста /Компьютерные технологии в управлении, диагностике и образовании (КТУДО-2002) //Сборник трудов международной научно-технической конференции. – Тверь, Тверской государственный технический университет, 2002. – С. 98 – 102.
3. Коломойцева И.А. Особенности применения существующих теорий «понимания» текста на естественном языке к медицинским текстам //Научные труды Донецкого государственного технического университета. Серия: Проблемы моделирования и автоматизации проектирования динамических систем. Вып. 29. – Севастополь: «Вебер», 2001. – С. 94–99.
4. Логический подход к искусственному интеллекту: От модальной логики к логике баз данных/ Пер. с франц./Тейз А., Грибомон П., Юлен Г. и др. - М.:Мир, 1998. – 494с.
5. Осипов Г. С. Приобретение знаний интеллектуальными системами: Основы теории и технологии. – М.: Наука. Физматлит, 1997. – 112 с.
6. Поспелов Д. А. Логико-лингвистические модели в системах управления. – М.: Энергоиздат, 1981. – 232 с.