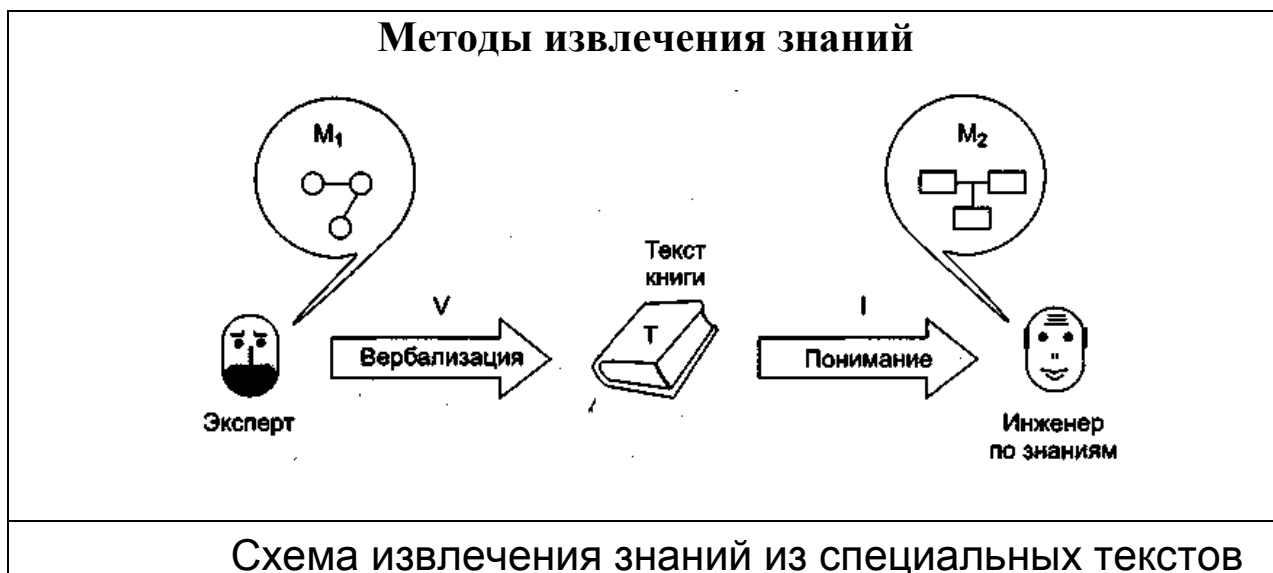


КУРС «Компьютерные технологии обучения»

Тема «Основы построения баз знаний (часть 2)»

Методы извлечения знаний

Группа *текстологических методов* объединяет методы извлечения знаний, основанные на изучении специальных текстов из учебников монографий, статей, методик и других носителей профессиональных знаний.



Методы извлечения знаний

Основными моментами **понимания текста** являются:

- Выдвижение предварительной **гипотезы** о смысле всего текста (предугадывание).
- Определение значений **непонятных слов** (то есть специальной терминологии).
- Возникновение общей гипотезы о содержании текста (**о знаниях**).
- Уточнение значения терминов и интерпретация отдельных фрагментов текста под влиянием общей гипотезы (**от целого к частям**).
- Формирование некоторой **смысловой структуры текста** за счет установления внутренних связей между отдельными важными (ключевыми) словами и фрагментами, а также за счет образования абстрактных понятий, обобщающих конкретные фрагменты знаний.
- Корректировка общей гипотезы относительно содержащихся в тексте фрагментов знаний (**от частей к целому**).
- Принятие основной гипотезы.

Методы извлечения знаний

Подготовкой к прочтению специальных текстов является выбор совместно с экспертами некоторого **«базового» списка литературы**, который постепенно введет аналитика в предметную область. В этом списке могут быть учебники для начинающих, главы и фрагменты из монографий, популярные издания. Только после ознакомления с **«базовым» списком** целесообразно приступать к чтению **специальных текстов**.

На процесс понимания влияют:

- экстракт компонентов почерпнутый из текста;
- предварительные знания аналитика о предметной области;
- общенаучная эрудиция аналитика;
- его личный опыт.

Методы извлечения знаний

Процедура **разбивки текста на части** («смысловые группы»), а затем сгущение, **сжатие содержимого каждого смыслового куска** в «смысловую вежу» является, видимо, основой для любого индивидуального процесса понимания. Такая компрессия (сжатие) текста в виде **набора ключевых слов**, передающих основное содержание текста, может служить удобной методологической **основой для проведения текстологических процедур извлечения знаний**.

Методы извлечения знаний

В качестве ключевого слова может служить **любая часть речи** (существительное, прилагательное, глагол и т. д.) или их **сочетание**. *Набор ключевых слов* (НКС) — это **набор опорных точек**, по которым разворачивается текст при кодировании» в память и осознается при декодировании, это **семантическое ядро цельности**.

Методы извлечения знаний

Одна из гипотез лингвостатистики:

Наиболее употребляемые слова являются наиболее важными с точки зрения содержания текста, то есть отражают его тематическую структуру.

Методы извлечения знаний

Три вида текстологических методов:

- анализ специальной литературы;
- анализ учебников;
- анализ методик.

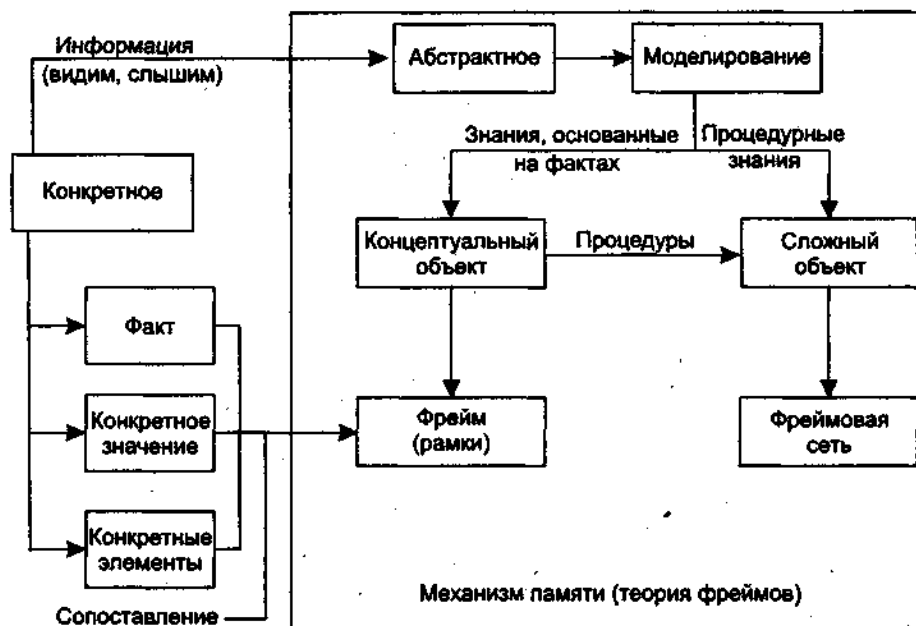
Методы извлечения знаний

Перечисленные три метода существенно отличаются по степени концентрированности специальных знаний, и по соотношению специальных и фоновых знаний.

Наиболее простым методом является анализ учебников, в которых логика изложения обычно соответствует логике предмета и поэтому макроструктура такого текста будет, наверное, более значима, чем структура текста какой-нибудь специальной статьи.

Анализ методик затруднен как раз сжатостью изложения и практическим отсутствием комментариев, то есть фоновых знаний, облегчающих понимание для неспециалистов. Поэтому можно рекомендовать для практической работы **комбинацию перечисленных методов.**

Методы извлечения знаний



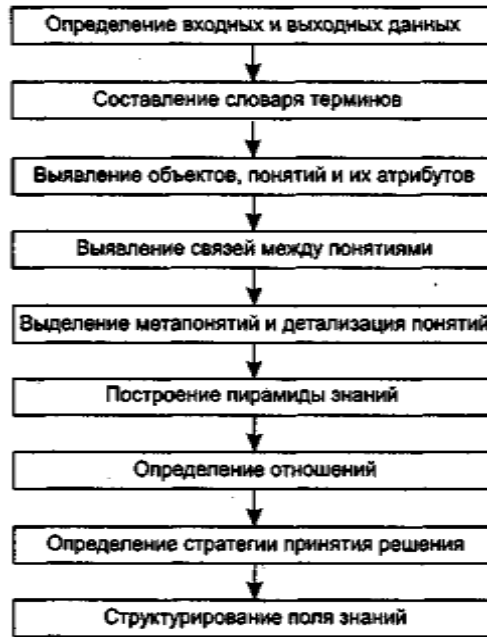
Методика анализа текстов с целью извлечения и структурирования знаний.

Методы извлечения знаний

Алгоритм извлечения знаний из текста

1. Составление «базового» списка литературы для ознакомления с предметной областью и чтение по списку.
2. Выбор текста для извлечения знаний.
3. Первое знакомство с текстом (беглое прочтение). Для определения значения неизвестных слов — консультации со специалистами или привлечение справочной литературы.
4. Формирование первой гипотезы о макроструктуре текста.
5. Внимательное прочтение текста с выписыванием ключевых слов и выражений, то есть выделение «смысловых вех» (компрессия текста).
6. Определение связей между ключевыми словами, разработка макроструктуры текста в форме графа или «сжатого» текста (реферата).
7. Формирование поля знаний на основании макроструктуры текста.

Простейшие методы структурирования знаний.



Стадии структурирования знаний

Простейшие методы структурирования знаний

1. Определение входных $\{X\}$ и выходных $\{Y\}$ данных. Этот шаг совершенно необходим, так как он определяет направление движения в поле знаний — от X (исходных данных) к Y (решению проблемы). Кроме того, **структура входных и выходных данных существенно влияет на форму и содержание поля знаний**. На этом шаге определение может быть достаточно размытым, в дальнейшем оно будет уточняться.

Простейшие методы структурирования знаний

2. Составление **словаря терминов** и наборов ключевых слов **N**. На этом шаге проводится текстуальный анализ всех протоколов сеансов извлечения знаний и выписываются все значимые слова, обозначающие понятия, явления, процессы, предметы, действия, признаки и т. п. При этом следует попытаться разобраться в значении терминов. Важен **осмысленный словарь**.

Простейшие методы структурирования знаний

3. Выявление объектов и понятий **{A}**. Производится «просеивание» словаря **N** и выбор **значимых для принятия решения понятий и их признаков**. В идеале на этом шаге образуется полный систематический набор терминов из какой-либо области знаний.

Простейшие методы структурирования знаний

4. **Выявление связей между понятиями.** Все в мире связано. Но определить, как направлены связи, что ближе, а что дальше, необходимо на этом этапе. Таким образом, строится сеть ассоциаций, где связи только намечены, но пока не поименованы. Например, понятия «день», «ночь», «утро» и «вечер» явно как-то связаны, связаны также и понятия «красный флаг» и «красный галстук», но характер связи тут существенно отличен.

Простейшие методы структурирования знаний

5. **Выявление метапонятий и детализации понятий.** Связи, полученные на предыдущем шаге, позволяют инженеру по знаниям структурировать понятия и как **выявлять понятия более высокого уровня обобщения** (метапонятия), так и **детализировать** на более низком уровне.

Простейшие методы структурирования знаний

6. Построение пирамиды знаний. Под пирамидой знаний мы понимаем иерархическую лестницу понятий, подъем по которой означает углубление понимания и повышения уровня абстракции (обобщенности) понятий. **Количество уровней в пирамиде зависит** от особенностей предметной области, профессионализма экспертов и инженеров по знаниям.

Простейшие методы структурирования знаний

7. Определение отношений. Отношения между понятиями выявляются как внутри каждого из уровней пирамиды, так и между уровнями. Фактически на этом шаге даются имена тем связям, которые обнаруживаются на шагах 4 и 5, а также обозначаются **причинно-следственные, лингвистические, временные и другие виды отношений.**

Простейшие методы структурирования знаний**8. Определение стратегий принятия решений (S_f).**

Определение стратегий принятия решения, то есть выявление цепочек рассуждений, связывает все сформированные ранее понятия и отношения в динамическую систему поля знаний. Именно стратегии придают активность знаниям, именно они «перетряхивают» модель M в поиске от X к Y .

Простейшие методы структурирования знаний

На практике при использовании данного алгоритма можно столкнуться с непредвиденными трудностями, связанными с ошибками на стадии извлечения знаний и с особенностями знаний различных предметных областей. Тогда возможно привлечение других, более «прицельных» методов структурирования. При этом на разных этапах схемы возможно использование различных методик.

Специальные методы структурирования

Используя представленный алгоритм, инженер по знаниям может испытывать необходимость в применении **специальных методов структурирования** на разных шагах алгоритма. При этом для таких простых и очевидных шагов, как определение входных и выходных понятий или составление словаря, никаких искусственных методов предлагаться не будет

Методы выявления объектов, понятий и их атрибутов

Понятие или **концепт** — это обобщение предметов некоторого класса по их специфическим признакам. связано с различными формами репрезентации понятий в памяти человека.

Все методы выявления понятий делятся на:

- *традиционные*, основанные на математическом аппарате распознавания образов и классификации;
- *нетрадиционные*, основанные на методологии инженерии знаний.

Методы выявления объектов, понятий и их атрибутов

Тридцати студентам, имеющим права на вождение автомобиля, предложили составить словарь терминов предметной области с помощью четырех методов:

1. Формирование перечня понятий (17 %).
2. Интервьюирование специалистов (35 %).
3. Составление списка элементарных действий (18 %).
4. Составление оглавления учебника (30 %).

Цифры в скобках характеризуют продуктивность соответствующего метода, то есть показывают, какой процент понятий из общего выявленного списка (702 термина) был получен соответствующим методом.

Методы выявления связей между понятиями

Концепты не существуют независимо, они включены в **общую понятийную структуру** с помощью *отношений*. Выявление связей между понятиями при разработке баз знаний доставляет инженеру по знаниям немало проблем. То, что знания в памяти — это некоторые связанные структуры, а не отдельные фрагменты, общеизвестно и очевидно. Тем не менее, основной упор в существующих моделях представления знаний делается на понятия, а связи вводят весьма примитивные (в основном причинно-следственные).

Методы выявления связей между понятиями

Все методы выявления связей можно разделить на две группы:

- Формальные.
- Неформальные (основаны на дополнительной работе с экспертом).

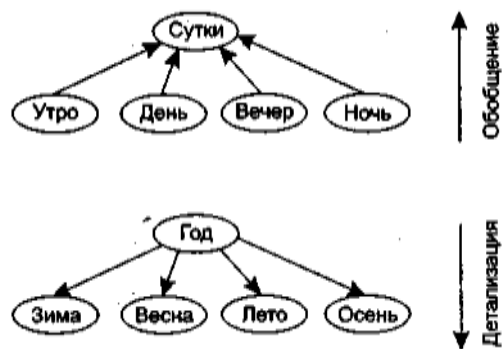
Неформальные методы выявления связей придумывает инженер по знаниям для того, чтобы вынудить эксперта указать явные и неявные связи между понятиями. Наиболее распространенным является метод «сортировка карточек» в группы, широко применяемый и для формирования понятий. Другим неформальным методом является построение замкнутых кривых. В этом случае эксперта просят обвести замкнутой кривой связанные друг с другом понятия. Этот метод может быть реализован как на бумаге, так и на экране дисплея. В этом случае можно говорить о привлечении элементов когнитивной графики.

После того как определены связи между понятиями, все понятия как бы распадаются на группы. Такого рода группы представляют собой метапонятия, присвоение имен которым происходит на следующей стадии процесса структурирования.

Методы выделения метапонятий и детализация понятий (пирамида знаний)

Процесс образования *метапонятий*, то есть интерпретации групп понятий, полученных на предыдущей стадии, как и обратная процедура — детализация (разукрупнение) понятий, — видимо, **принципиально не поддающиеся формализации операции.**

Методы выделения метапонятий и детализация понятий



. Обобщение и детализация понятий

Методы определения отношений

В работе [Поспелов, 1986] указывается на наличие более 200 базовых видов различных отношений, существующих между понятиями. Предложены различные классификации отношений [Келасьев, 1984; Поспелов, 1986]. Следует только подчеркнуть, что помимо универсальных отношений (пространственных, временных, причинно-следственных) существуют еще и специфические отношения, присущие той или иной предметной области [Гаврилова, Червинская, Яшин, 1988].

Методы определения отношений

Интересные возможности к структурированию знаний добавляют системы когнитивной графики. Так, в системе OPAL [Olton., Muser, Combs et al., 1987] эксперт может манипулировать на экране дисплея изображениями простейших понятий и строить схемы лечения заболеваний, обозначая отношения явными линиями, которые затем именуется.

Методы определения отношений

Скудность методов структурирования объясняется тем, что методологическая база инженерии знаний только закладывается, а большинство инженеров по знаниям проводит концептуализацию, руководствуясь наиболее дорогими и неэффективными способами — «проб и ошибок» и «по наитию», то есть исходя из соображений здравого смысла.