Лекция № 8.

Методы извлечения знаний

Методы извлечения знаний

1. Классификация методов извлечения знаний

Основной принцип деления методов извлечения знаний (МИЗ) связан с источником знаний.

Коммуникативные методы охватывают все виды контактов с живым источником знаний – экспертом.

Текстологические методы предполагают извлечение знаний из документов (методик, пособий, руководств) и специальной литературы (статьи, монографии, учебники).

Разделение этих групп методов не означает их антагонистичности – обычно ИЗ комбинирует различные методы: сначала изучает литературу, а затем идет беседовать с экспертом и наоборот.

Классификация методов извлечения знаний

Коммуникативные методы делятся на *активные* и *пассивные*.

В **пассивных методах** ведущую роль играет эксперт, а инженер знаний только протоколирует рассуждения эксперта во время его реальной работы по принятию решений или записывает то, что эксперт считает нужным сказать во время беседы или лекции. Пассивные методы на 1-й взгляд просты, но они требует от ИЗ умения четко анализировать «поток сознания Э и выявлять в нем значимые фрагменты знаний.

В **активных методах** ведущая роль, инициатива принадлежит ИЗ, который активно контактирует с Экспертом различными способами – в играх, диалогах и т.п.

Активные и пассивные методы могут чередоваться даже в рамках одного сеанса извлечения знаний.

Классификация методов извлечения знаний

Активные методы делятся на две группы в зависимости от числа участвующих экспертов – *индивидуальные* и *групповые*.

Групповые методы служат для активация мышления участников дискуссий и позволяют выявлять весьма нетривиальные аспекты их знаний

Индивидуальные методы все же остаются ведущими, поскольку столько деликатная процедура, как «отъём знаний» не терпит свидетелей.

2. Выбор метода извлечения знаний

На выбор метода извлечения знаний влияют 3 фактора:

- *личностные особенности инженера знаний,*
- *личностные особенности эксперта,*
- *характер предметной области*.

По психологическим характеристикам люди делятся на 3 типа:

1. *Мыслитель* (познавательный тип).
2. *Собеседник* (эмоционально-познавательный тип).
3. *Практик* (практический тип).

Мыслители ориентированы на интеллектуальную работу, учебу, теоретические обобщения.

Собеседники – это общительные, открытые люди, готовые к сотрудничеству.

Практики предпочитают разговорам действие, хорошо реализуют замыслы других, нацелены на результативную работу.

Выбор метода извлечения знаний

Для характеристики предметной области используют такое ее качество как *структурированность*, т.е. степень теоретического осмысления и выявления основных закономерностей и принципов, действующих в данной ПО.

Хорошо структурированная ПО – область с четкой аксиоматизацией, широко применяется математический аппарат, теоретически устоявшаяся.

Средне структурированная ПО – область с определившейся терминологией, явными взаимосвязями между явлениями, но развивающейся теорией.

Слабо структурированная ПО – область с размытыми определениями, скрытыми взаимосвязями между явлениями, богатой эмпирикой, с большим количеством «белых пятен».

3. Пассивные методы извлечения знаний

Пассивные методы делятся на *наблюдения, вербальные отчеты, лекции*.

1. Наблюдение.

Наблюдение обычно используется на начальных стадиях и применяются не самостоятельно, а в сочетании с другими методами.

Суть: инженер знаний находится непосредственно рядом с экспертом во время его профессиональной деятельности или имитации такой. При этом ИЗ заранее просит Э максимально полно комментировать все его действия. ИЗ записывает все действия Э, его реплики, объяснения (возможна видеозапись сеанса).

Непременное условие – невмешательство ИЗ в работу Э. Поэтому именно наблюдение – единственно «чистый» метод, исключающий вмешательство ИЗ и навязывание им каких-то своих представлений.

Пассивные методы извлечения знаний

Существует две разновидности проведения наблюдений:
- наблюдение за реальным процессом,
- наблюдение за имитацией процесса.

Обычно используют обе разновидности.
Очень важно видеть реальный процесс для *разработки пользовательского интерфейса*, ведь ЭС будет работать в условиях реального производственного процесса.

Сеансы наблюдений требуют от инженера знаний:
- овладения техникой стенографии.
- знаний методик хронометрирования производственного процесса.
- навыков «чтения по глазам»,
- серьезного предварительного знакомства с ПО.

Протоколы наблюдений тщательно *расшифровываются* и затем *обсуждаются* с экспертом.

Пассивные методы извлечения знаний

2. Вербальные отчеты.

Суть: в отличие от наблюдений здесь ИЗ просит Э не просто комментировать свои действия и решения, но и объяснять, как это решение было найдено, т.е. продемонстрировать всю цепочку рассуждений. Все это протоколируется ИЗ. Использование магнитофонов и диктофонов нежелательно – они сковывают Э.

Основная сложность метода: принципиальная сложность для любого человека *объяснить, как он думает*. Люди не всегда способны достоверно описывать свои мыслительные процессы. Вообще часть знаний у человека существует в невербальной форме, не поддается словесному описанию (различные процедурные знания типа «как завязывать шнурки».

Хотя есть люди (эксперты), склонные к рефлексии, для которых эта работы вполне доступна.

Пассивные методы извлечения знаний

Удачное протоколирование «мыслей вслух»: делает этот метод наиболее эффективным методом извлечения знаний, поскольку Э может проявить себя максимально ярко и полно.
Здесь Э может блеснуть свой эрудицией, показать всю глубину своих знаний.
Этот метод – наиболее лестный и приятный для эксперта.

Этот метод требует от ИЗ тех же навыков, что и *наблюдение*.
Также требуется расшифровка протоколов «мыслей вслух» - вербальных отчетов и их обсуждение с экспертом.

Пассивные методы извлечения знаний

3. Лекции.

Лекция – самый старый и известный способ передачи знаний.
ИЗ должен уметь слушать лекции, так как он не может учить Э, как нужно читать лекции.

ИЗ должен сформулировать для Э тему цикла лекций, а также тему и задачу каждой лекции (чтобы ограничить степень его свободы). Например, тема цикла лекций «Постановка диагноза – воспаление легких». Тема конкретной лекции «Анализ рентгенограмм», задача лекции – научить слушателей по перечисленным Э признакам ставить диагноз «воспаление легких» и делать прогноз.

При такой постановке Э может заранее структурировать свои знания и ход рассуждений. ИЗ нужно только все внимательно записать Писать конспект – это искусство: нужно записывать главное, опускать второстепенное, выделять фрагменты знаний (параграфы, подпараграфы), уметь обобщать.

Пассивные методы извлечения знаний

В конце или по ходу лекции необходимо задавать уточняющие вопросы.
Хороший вопрос по ходу лекции помогает и лектору и слушателю.

Опытный лектор разбивает все вопросы условно на 3 группы:
- Умные вопросы, углубляющие лекцию,
- Глупые вопросы или вопросы «не по существу»,
- Вопросы «на засыпку» или провокационные вопросы.

Цикл (курс) лекций обычно включает 3-5 лекций.
Продолжительность лекции 40-50 мин, с 5-10-минутным перерывом.
Этот метод, как и первые два используется на начальной стадии разработки БЗ ЭС для эффективного погружения в ПО.

4. Активные индивидуальные методы

Активные индивидуальные методы: *анкетирование, интервью, свободный диалог, экспертные игры*.

Во всех этих методах активную роль играет ИЗ (аналитик), который пишет сценарий и режиссирует сеансы извлечения знаний.
Первые три метода, которые можно назвать вопросными методами поиска знаний, схожи между собой и различаются лишь степенью свободы, которую может себе позволить ИЗ при проведении сеансов извлечения знаний. Игры же имеют ряд существенных отличий.

Активные индивидуальные методы

1. Анкетирование.

Анкетирование — наиболее жесткий метод, поскольку он наиболее стандартизирован. В этом случае ИЗ заранее составляет вопросник или анкету, размножает ее и использует для опроса нескольких экспертов.

Анкетирование может проводиться двумя способами:

1) ИЗ вслух задает вопросы и сам заполняет анкету по ответам эксперта.
2) Эксперт самостоятельно заполняет анкету после предварительного инструктирования.

Выбор способа зависит от конкретных условий (например, от оформления анкеты, ее понятности, готовности специалистов отвечать на вопросы).

Второй способ кажется более предпочтительным, так как эксперт не ограничен временем на обдумывание ответов.

Активные индивидуальные методы

Анкета заслуживает особого внимания. Существует несколько общих рекомендаций при составлении анкет:

1. Анкета не должна быть монотонной и однообразной, т.е. вызывать скуку или усталость. Это достигается вариациями формы вопросов, сменой тематики, вставкой вопросов-шуток и игровых вопросов.
2. Анкету следует приспосабливать к языку эксперта.
3. Вопросы влияют друг на друга, поэтому их последовательность должна быть строго продуманной.
4. Желательно стремиться к оптимальной избыточности. Известно, что в анкете всегда много лишних вопросов; часть из них необходима — это так называемые контрольные вопросы, а другая часть должна быть минимизирована.
5. Анкете необходимы «хорошие манеры», т.е. ее язык должен быть ясным, понятным, предельно вежливым.

Активные индивидуальные методы

2. Интервью.

Под интервью будем понимать специфическую форму общения ИЗ и Э, в которой ИЗ задает серию заранее подготовленных вопросов с целью извлечения знаний о предметной области.

Интервью близко тому способу анкетирования, когда ИЗ сам заполняет анкету, занося в нее ответы специалистов.

Основное преимущество интервью в гибкости; в этих условиях ИЗ может опускать ряд вопросов в зависимости от ситуации, вставлять новые, изменять темп, разнообразить ситуацию общения.

Кроме того, у ИЗ появляется возможность «взять в плен» Э своим обаянием, заинтересовать его самой процедурой и тем самым увеличить ее эффективность.

Активные индивидуальные методы

Есть три основные *характеристики вопросов*, которые влияют на *качество интервью*. Это:

1) язык вопроса (понятность, лаконичность, терминология);
2) порядок вопросов (логическая последовательность и немонотонность);
3) уместность вопросов (этика, вежливость).

Вопрос в интервью — это не просто средство общения, но и способ передачи мыслей и позиции аналитика. Отсюда необходимость в протоколах фиксировать не только ответы, но и вопросы, предварительно тщательно отрабатывая их форму и содержание.

Вопросы имеют для Э диагностическое значение — несколько откровенно глупых вопросов могут полностью разочаровать Э и отбить у него охоту к дальнейшему сотрудничеству.

Известен ответ Маркса на вопрос Прудона: «Вопрос был до такой степени неправильно поставлен, что на него невозможно было дать правильный ответ».

Активные индивидуальные методы

Виды вопросов

Все вопросительные предложения можно разбить на два типа.

1. **Вопросы с неопределенностью**, относящиеся ко всему предложению (например: «Действительно введение антибиотиков в больших дозах может вызвать шок»?).
2. **Вопросы с неполной информацией** (например: «При каких условиях Вас не устраивает этот препарат?»), часто начинающиеся со слов «кто», что», «где», «когда» и т.д.

Открытый вопрос называет тему или предмет, оставляя экспертам полную свободу в том, что касается формы и содержания ответа.

В **закрытом** вопросе эксперт выбирает ответ из набора предложенных.

Личный вопрос касается непосредственно индивидуального опыта специалиста («Что бы Вы порекомендовали принимать при гриппе?»). Личные вопросы обычно активизируют мышление специалиста, «играют» на его самолюбии; они всегда украшают интервью.

Безличный вопрос направлен на выявление наиболее распространенных и общепринятых закономерностей предметной области («Что влияет на заболеваемость?»).

Активные индивидуальные методы

Если эксперт «закрепощен», возможно использование не **прямых вопросов**, которые непосредственно указывают на предмет или тему («Применяете ли Вы методы сетевого планирования?»), а **косвенных**, которые лишь неявно указывают на интересующий предмет («А что Вы думаете о целесообразности применения…»).

Вербальные вопросы — это традиционные устные вопросы.

Вопросы с использованием наглядного материала разнообразят интервью и снижают утомляемость интервьюируемого. В таких вопросах используют фотографии, рисунки и карточки. Например, специалисту предлагаются цветные картонные карточки, на которых выписаны признаки какого-нибудь заболевания. Затем ИЗ просит разложить эти карточки в порядке убывания значимости признаков.

Контрольные вопросы применяют для проверки достоверности и объективности информации, полученной в интервью ранее.

Контрольные вопросы должны быть составлены хитро, чтобы не обидеть эксперта недоверием (для этого используют повторение вопросов в другой форме, уточнения, ссылки на другие источники).

Активные индивидуальные методы

3. Свободный диалог

Свободный диалог — это метод извлечения знаний в форме беседы ИЗ и Э, в которой нет жесткого регламентированного плана и вопросника.
Это определение не означает, что к свободному диалогу не надо готовиться. Напротив, внешне свободная и легкая форма требует серьезной профессиональной и психологической подготовки.
Подготовка занимает разное время в зависимости от степени профессионализма ИЗ (аналитика), но в любом случае она необходима, так как несколько уменьшает вероятность самого нерационального метода — проб и ошибок.

Подготовка к диалогу включает составление плана проведения сеанса, в котором необходимо предусмотреть следующие стадии:

1. *Начало беседы* (знакомство, создание у эксперта «образа» аналитика, объяснение целей и задач работы).
2. *Диалог по извлечению знаний*.
3. *Заключительная стадия* (благодарность за потраченное время, подведение итогов, договоренность о последующих встречах).

Активные индивидуальные методы

Квалифицированная подготовка к диалогу помогает аналитику стать истинным драматургом, т.е. запланировать гладкое течение процедуры извлечения знаний: приятное впечатление в начале беседы с переходом к профессиональному контакту через пробуждение интереса и завоевание доверия эксперта.

Для поддержания желания специалиста продолжать беседу полезны бывают «поглаживания» (в терминологии Э. Берна), т.е. подбадривающие собеседника фразы-вставки: «Я Вас понимаю...», «...это очень интересно» и т. п.

В свободном диалоге важно выбрать правильный темп или ритм беседы: без больших пауз, так как специалист может отвлечься, но и без «гонки», иначе оба участника быстро утомляются и нарастает напряженность; кроме того, некоторые люди говорят и думают медленно.

Умение чередовать разные темпы, напряжение и разрядку в беседе существенно влияет на результат.

Активные индивидуальные методы

4. Экспертные игры

Игра — вид человеческой деятельности, который отражает (воссоздает) другие ее виды. При этом для их характерна одновременно и условность и серьезность.

Понятие экспертной игры восходит к 3-м источникам – деловой игры, диагностической игры и компьютерной игры, применяемой в обучении.

Деловая игра – эксперимент, где участникам предлагается производственная ситуация, а они на основе своих опыта и знаний принимают решения. Принятые решения анализируются и вскрываются закономерности мышления участников игры. Именно эта – анализирующая часть игры и полезна для извлечения знаний. Если участниками деловой игры являются эксперты, то сама играет становится экспертной.

Активные индивидуальные методы

Диагностическая игра – это та же деловая игра, но применяемая конкретно для диагностики принятия решений в медицине. Эти игры возникли при исследовании передачи опыта от опытных врачей к новичкам.

Экспертная игра или *игра с экспертом.*
Здесь с экспертом играет инженер знаний который берет на себя какую-нибудь роль в моделируемой ситуации.
Например, в игре «Учитель-ученик» ИЗ берет на себя роль ученика, а эксперт – учителя и поправляет ошибки ученика при выполнении роли эксперта в какой-нибудь ПО.
Возможно и такое распределение ролей:
«врач» (инженер знаний) – «консультант» (эксперт).

В экспертной игре ИЗ может проверить свои гипотезы о ПО, получить новые знания.

5. Активные групповые методы

Основное достоинство групповых методов - это возможность одновременного «поглощения» знаний от нескольких специалистов (экспертов), взаимодействие которых вносит в этот процесс элемент принципиальной новизны: индивидуальное или персональное знание, разные взгляды и позиции существенно обогащают общее поле знаний. Однако следует отметить, что эти методы гораздо более трудоемки и дороги, чем индивидуальные, по причине сложности их организации.

Активные групповые методы обычно используются как своего рода острая приправа на стадии извлечения знаний, сами по себе они не могут служить источником более или менее полного знания. Их применяют как дополнение к традиционным индивидуальным методам (наблюдения, интервью и т. д.), ради активизации мышления и поведения экспертов.

Активные групповые методы:

«круглый стол»,
«мозговой штурм»,
ролевые игры.

Активные групповые методы

1. «Круглый стол».

Метод круглого стола (термин заимствован из журналистики) предусматривает обсуждение какой-либо проблемы из выбранной предметной области, в котором принимают участие с равными правами несколько специалистов.

Обычно вначале дискуссии участники высказываются в определенном порядке, а затем переходят к живой свободной дискуссии.

Число участников дискуссии колеблется от 3 до 5-7.

Задача дискуссии - коллективно, с разных точек зрения, под разными углами обсудить и исследовать спорные моменты и гипотезы в определенной ПО.

Для остроты дискуссии на круглый стол приглашают представителей разных «научных школ» и разных поколений. Это уменьшает опасность получения односторонних однобоких знаний.

Ход беседы за круглым записывается на магнитофон, а затем анализируется.

Активные групповые методы

Перед началом дискуссии ведущему полезно:
- убедиться, что все правильно понимают цель дискуссии (т. е. что происходит сеанс извлечения данных и знаний);
- установить регламент выступлений (не более 5-7 мин);
- четко сформулировать тему дискуссии.

По ходу дискуссии важно проследить, чтобы слишком эмоциональные и разговорчивые специалисты не подменили тему и чтобы критика позиций друг друга была обоснованной.

Активные групповые методы

2. «Мозговой штурм».

«Мозговой штурм» - один из наиболее распространенных методов раскрепощения и активизации мышления.

Впервые этот метод использовал в 1939 году в США А. Осборн как способ получения новых идей в условиях запрещения критики.

Замечено, что боязнь критики мешает творческому мышлению, поэтому ***основная идея мозгового штурма*** - это отделение процедуры генерирования идей в замкнутой группе специалистов от процесса анализа и оценки высказанных идей.

Основной девиз штурма - «Чем больше идей, тем лучше».

Активные групповые методы

Как правило, штурм длится недолго (около 40 мин). Участникам (до 10 человек) предлагается высказывать любые идеи (шутливые, фантастические, ошибочные) на заданную тему (критика запрещена). Обычно высказывается более 50 идей.

Регламент выступлений - до 2 мин.

Самый интересный момент штурма - это наступление пика (ажиотажа), когда идеи начинают «фонтанировать», т. е. происходит непроизвольная генерация гипотез участниками.

При последующем анализе всего лишь 10-15% идей оказываются разумными, но среди них бывают весьма оригинальные.

Оценивает результаты обычно группа экспертов, не участвовавшая в генерации идей.

Активные групповые методы

Ведущий «мозгового штурма» - аналитик (ИЗ) - должен свободно владеть аудиторией, подобрать активную группу специалистов - «генераторов», не зажимать плохие идеи - они могут служить катализаторами хороших. Искусство ведущего - это искусство задавать вопросы аудитории, «подогревая» генерацию. Вопросы служат «крючком», при помощи которого извлекаются идеи.

Вопросы также могут останавливать излишне многословных экспертов и служить развитию идей других.

Фиксация хода сеанса - традиционная (протокол или магнитофон).

Активные групповые методы

3. «Ролевые игры».

Играют одновременно несколько экспертов.
Заранее составляется сценарий, распределяются роли, к каждой роли готовится портрет-описание и разрабатывается система оценивания игроков.

В игре принимает участие от 3 до 6 человек. Если участников больше, то они разбиваются на конкурирующие бригады. Вводится элемент состязательности, например, «чей диагноз лучше», «кто быстрее найдет неисправность» и т.п.
Возможно использование наглядных материалов, табличек («Директор», «Главный конструктор», ..).
Но главное, чтобы эксперты в игре действительно заиграли, раскрепостились и «раскрыли свои карты».

6. Текстологические методы

Текстологические методы – объединяют группу методов извлечения знаний, основанных на изучении специальных текстов из учебников, монографий, статей, методик и других носителей знаний.

Задачу извлечения знаний из текстов можно сформулировать как задачу **понимания** и выделения **смысла** текста.
Понимание, здесь, *это формирование семантической или понятийной структуры текста.*
Необходимо выделять *эксплицитные* (явные) и *имплицитные* (неявные) связи в тексте.
Эксплицитные связи отражены в структуре текста (разбиение на главы и параграфы).
Имплицитные связи – это внутренние связи между «смысловыми структурами», которые представляются наборами ключевых слов.
Сжатие текста в виде набора ключевых слов, передающего основное содержание текста, является удобной методологической основной для извлечения знаний.

Текстологические методы

Методика анализа текстов с целью извлечения и структурирования знаний:

1. Составление базового списка литературы для ознакомления с ПО.
2. Выбор текста для извлечения знаний.
3. Беглое чтение для определения значений незнакомых слов – консультации со специалистами или привлечение справочной литературы.
4. Формирование 1-й гипотезы о макроструктуре текста.
5. Внимательное прочтение текста с выписыванием ключевых слов и выражений.
6. Определение связей между ключевыми словами, разработка макроструктуры текста в виде графа и «сжатого» текста (реферата).
7. Формирование концептуального описания – основных взаимосвязей между понятиями ПО.