

ПСИХОЛОГИЧЕСКИЙ ПОДХОД К ИЗВЛЕЧЕНИЮ ЗНАНИЙ В ПРОЦЕССЕ ЧЕЛОВЕКО-КОМПЬЮТЕРНОГО ВЗАИМОДЕЙСТВИЯ С СИСТЕМОЙ ЭЛЕКТРОННОГО ДОКУМЕНТООБОРОТА

Арсений Баканов

Аннотация: В настоящей статье описывается способ извлечения и визуализации экспертных знаний в процессе человеко-компьютерного взаимодействия с системами электронного документооборота, системами документационного обеспечения управления. В работе приведено описание проведенных экспериментов, приводятся полученные результаты экспериментов. Описываемый подход к извлечению знаний, ориентирован на использование в системах документационного обеспечения управления и предполагает использование алгоритмов и правил для выявления эвристик в процессе принятия решений при решении управленческих задач.

Ключевые слова: Системы электронного документооборота, Извлечение экспертных знаний, Представление знаний.

Классификация ключевых слов ACM: H.4.2 Types of Systems - Decision support

Введение

Статья посвящена исследованию проблем, связанных с взаимодействием человека с информационными системами, в частности вопросам извлечения и визуализации экспертных знаний в процессе взаимодействия с системами документационного обеспечения управления. Важность и актуальность проводимых исследований обусловлена рядом причин, среди которых в первую очередь необходимо отметить востребованность в связи, с созданием в РФ электронного правительства и переходу к безбумажному документообороту в Федеральных органах исполнительной власти РФ. В настоящей статье описывается оригинальная методика неявного извлечения экспертных знаний в процессе человеко-компьютерного взаимодействия. Методика ориентирована на использование в системах документационного обеспечения управления и предполагает разработку алгоритмов и правил для выявления эвристик в процессе принятия решений при решении управленческих задач. Также в данной статье предлагается способ визуализации извлеченных экспертных знаний, позволяющий представить извлеченные знания как структурированную совокупность символов, образов и массивов текстовой информации.

В настоящей статье, кратко представлены результаты исследований, проводимых с 2012 года по настоящее время в Институте психологии РАН. Проводились исследования двух типов: в исследованиях первого типа первоначально предъявлялся стимульный материал (в виде вопросов, ключевых слов, графических образов), а затем испытуемому предъявлялась текстовая информация, после прочтения которой, испытуемый принимал решение о сортировке (т. е. определял департамент или подразделение организации, куда документ должен быть направлен для исполнения), отвечал на вопросы по тексту и проходил психологическое тестирование; в исследованиях второго типа, эксперимент начинался с

предъявления текстовой информации, после прочтения которой, испытуемый принимал решение о сортировке, отвечал на вопросы по тексту и проходил психологическое тестирование.

Испытуемым предъявлялись тексты, как прошедшие предварительную обработку – т.е. с выделенными опорными словами и визуализированной структурой, так и тексты без предварительной обработки.

В ходе исследований, использовалась установка контроля движения взора испытуемого (www.smivision.com). С ее помощью были получены экспериментальные данные: траектория взора, диаметр зрачка испытуемого, скорость перемещения взора, как в процессе чтения текста, так и в процессе принятия решения.

Системы электронного документооборота

Основной задачей системы электронного документооборота (СЭД) — является обеспечение процесса создания, управления и распространения электронных документов. Первоначально системы электронного документооборота использовались исключительно для учета и контроля прохождения документов, однако со временем к СЭД стали предъявляться все новые и новые требования, в связи с этим СЭД стали охватывать все более широкий спектр задач. В настоящее время СЭД фактически представляют собой системы организационного управления.

Информационные системы организационного управления в процессе функционирования накапливают большие информационные массивы [Баканова, 2007], сохраняемые в виде баз данных. В информационных массивах содержится информация, отражающая различные аспекты деятельности организации. Одним из таких аспектов деятельности организации является процесс принятия управленческих решений. Поэтому расширение возможностей использования информационного фонда систем организационного управления, за счет создания информационной поддержки принятия решений, является одной из актуальных задач повышения эффективности СЭД.

Обобщение и анализ информации, относящейся к различным прикладным задачам, открывает более широкие возможности ее использования. Исследование этих возможностей для целей, связанных с реализацией поддержки принятия решений в СЭД, представляет собой интересные малоизученные задачи. Одной из таких задач является выявление информационных источников для реализации сервисов поддержки управленческой деятельности. Ее решение позволит перейти к определению сервисов, необходимых для поддержки управленческой деятельности [Баканова, 2007]. Достоинством созданных сервисов будет оперативная и актуальная информация, полученная из выявленных источников данных. Анализ дополнительных возможностей использования накопленного информационного фонда и создание на его основе сервисов поддержки управленческой деятельности являются важной составной частью задачи повышения эффективности использования СЭД, а также качества управления в организационных структурах.

Для решения указанной задачи, при проектировании СЭД требуется выявить потенциал прикладных подсистем в части накопления данных, полезных для использования в управленческой деятельности. При этом, возможно, потребуется расширить функциональные задачи подсистем [Ташев и др., 2013] для формирования и сохранения информации, которая может быть использована в процессах принятия решений. Использование прикладных (технологических) информационных подсистем в качестве источника данных для организации поддержки принятия управленческих решений определяет новые подходы к разработке СЭД с целью создания востребованных сервисов поддержки управленческой деятельности. Функционал сервисов базируется на алгоритмах, обеспечивающих лицо, принимающее решение (ЛПР), оперативно подготовленной информацией, а также на методах выявления и

идентификации требуемых данных, содержащихся в программных комплексах систем организационного управления. Для эффективного функционирования системы поддержки принятия решений (ППР) требуется знание предметной области [Luger, 2002]. В целях успешной реализации системы, было принято решение ограничить предметную область транспортной тематикой. Такое решение было принято по примеру разработчиков создавших приложение SHRDLU [Winograd, 1972] работающее в минимальной предметной области.

Взаимодействие человека с системами электронного документооборота

В организационном управлении субъектом управления всегда выступает человек. Для принятия управленческих решений, направленных на выполнение стратегических и тактических задач, поставленных и оперативно возникающих перед организацией, руководитель постоянно должен принимать и отвергать сотни разнообразных альтернатив [Баканова, 2007]. Управленческое решение – это некоторое воздействие на управляемый объект, определяющее действия, необходимые для проведения изменений в его состоянии.

Проблемам принятия решений посвящено множество научных работ, среди которых можно отметить работы таких ученых как Luger, Newell, Simon, Ларичев, Петровский. Среди исследований, касающихся вопросов информационного взаимодействия человека с техникой вообще, необходимо отметить работу коллектива авторов: Т. В. Атанасовой, Т.Н. Савченко, Г.М. Головиной и др., в которой описаны психологические механизмы взаимодействия человека с интеллектуальной информационной средой обитания [Атанасова Т.В. и др., 2010]. С позиций когнитивной психологии [Величковский Б.М., 2006], возможности человека по приему и переработке информации описываются с помощью различных функциональных моделей структуры памяти пользователя [Baddeley, 2009].

В процессе взаимодействия с информационной средой, к человеку по различным информационным каналам (с использованием различных телекоммуникационных устройств) могут поступать значительные объемы информации для переработки и последующего принятия решения. Количество каналов информации, по которым информация поступает к человеку, продолжает стремительно увеличиваться, как и количество поступающей информации. В процессе взаимодействия с СЭД, человеку приходится учитывать значительное количество различных факторов, а также решать задачи многокритериального выбора. Для человеческой системы переработки информации многокритериальные задачи представляют собой особо сложный класс задач [Петровский А.Б., 2004]. Наличие многих критериев приводит к нагрузке на человеческую систему переработки информации, заставляя человека использовать различные, зачастую оригинальные эвристики для того, чтобы решить поставленную задачу [Ларичев О. И., 1987].

Описание проведенных экспериментов

В соответствии с целью работы было проведено экспериментальное исследование взаимодействия человека с интеллектуальной информационной средой на примере взаимодействия человека с системой поддержки принятия решений. В ходе исследований моделировалась работа лица принимающего решение. Работа ЛПР заключалась в ознакомлении с документом – чтении текста предъявленного на мониторе компьютера, с последующим принятием решения о сортировке (decision making), т. е. определении департамента или подразделения организации, куда документ должен быть направлен для дальнейшей обработки.

Необходимо отметить, что в крупные управленческие организации ежедневно поступают сотни документов и ошибка ЛПР при сортировке чревата не только проволочками при исполнении документа, но и ненадлежащим исполнением, в случае если документ будет направлен в непрофильное подразделение. Для работы с документами (в том числе входящими), в крупных управленческих организациях традиционно используют системы электронного документооборота [Баканова Н. Б., 2007]. В состав такой системы электронного документооборота может быть включен программный модуль, реализующий функции системы поддержки принятия решений. Данный модуль осуществляет предварительную обработку текста документа, выделяет опорные слова цветом и визуализирует структуру документа. В рамках проводимых экспериментов представляло научный интерес исследовать процесс принятия решений в задаче о сортировке документа, как «экспертами», т. е. людьми, имеющими опыт работы в качестве ЛПР, так и «новичками», т. е. людьми, не имеющими опыта работы в качестве ЛПР. В процессе исследования, решения принимались как с помощью системы поддержки принятия решений, так и без помощи системы поддержки принятия решений. В качестве испытуемых в исследованиях участвовали:

- 1) Эксперты – лица имеющие опыт работы в качестве ЛПР.
- 2) Новички – лица, не имеющие опыта работы в качестве ЛПР – студенты высших учебных заведений г. Москвы.

Исследования проводились с использованием оборудования отслеживающего траекторию движения глаз пользователя в процессе считывания информации, а также в процессе принятия решений (www.smivision.com). Параллельно с проведением экспериментов, проводились исследования когнитивного стиля импульсивный/рефлексивный [Холодная, 2002], а также стиля саморегуляции поведения (ССПМ) по методике В. И. Моросановой [Моросанова, Индина, 2011]. Испытуемым предъявлялся некоторый документ содержащий текст, после прочтения, которого необходимо было принять решение о выборе из некоторого числа альтернатив (принять решение о сортировке документа). Испытуемым последовательно предъявлялись тексты как прошедшие предварительную обработку системой поддержки принятия решений, так и без обработки. Обработка заключалась в определении структуры текста и проведении контент-анализа с использованием специально разработанного словаря (тезауруса). В результате обработки визуализировалась структура текста, а опорные слова выделялись цветом [Баканов А.С., 2009]. Последовательность предъявления текстов (с обработкой и без обработки) на мониторе компьютера, менялась, чтобы нивелировать привыкание испытуемого, после прочтения текста испытуемый принимал решение. Экспериментатор наблюдал и фиксировал (в том числе и в формате *.mpeg) траекторию движения глаз испытуемого, как в процессе чтения текста, так и в процессе принятия решения. На этапе принятия решения испытуемым предъявлялись (также на мониторе компьютера) вопросы двух типов. К вопросам первого типа относились вопросы, предъявляемые в следующем порядке:

- 1) Выберите альтернативы (из некоторого списка) к которым относится данный документ/текст.
- 2) Выберите альтернативы (из числа выбранных ранее) к которым данный документ/текст относится в наибольшей степени.
- 3) Выберите только одну альтернативу (из выбранных ранее).

Таким образом, в процессе ответов на вопросы первого типа в приведенном выше порядке выявлялась структура ментальных репрезентаций пользователя, сформировавшаяся по прочтении конкретного документа/текста см. Рисунок 1.

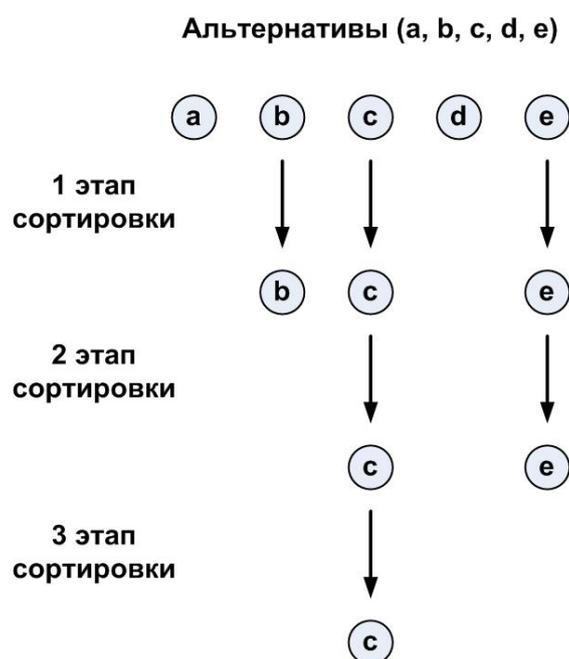


Рис. 1

Необходимо отметить, что существует достаточное количество определений термина ментальная репрезентация, мы же будем использовать следующее определение: ментальная репрезентация - субъективный образ объективной реальности, отражение внутреннего и внешнего мира в сознании человека. Или применительно к данному исследованию субъективный образ прочтенного документа/текста в сознании испытуемого. При ответе на вопрос (также предъявляемый на экране монитора) второго типа испытуемый должен был количественно оценить степень, в которой данный документ/текст относится к выбранной альтернативе. Траектория движения взора, скорость движения взора, а также диаметр зрачка испытуемого в процессе принятия решения фиксировалась экспериментатором.

Кроме траектории взора, экспериментатор в процессе проводимого исследования фиксировал время в течении которого испытуемый читал предъявленный текст, а также время в течении которого испытуемый принимал решение после прочтения текста как с выделенными цветом (опорными) словами, так и без выделенных цветом опорных слов. В настоящее время исследования продолжаются. К настоящему времени в исследованиях приняло участие более 70 человек.

В ходе проведения экспериментов удалось выявить, что большинство испытуемых читали предлагаемый текст дважды. Причем во второй раз взор испытуемого перемещался от одного опорного слова (фрагмента текста) к другому опорному слову (фрагменту текста) в независимости были ли они выделены в процессе обработки текста или нет. В процессе движения от одного опорного слова к другому, взор испытуемого „перескакивал” через строки и абзацы, иногда возвращаясь к отдельным фрагментам текста. Можно предположить, что таким образом испытуемый пытался запомнить наиболее важные (для него) слова или фрагменты текста. Такие последовательности слов можно представить в виде:

$$A(X_1, X_2, X_3, \dots, X_n)$$

$$B(Y_1, Y_2, Y_3, \dots, Y_m)$$

и т.д.

где $X_1, X_2, X_3, \dots, X_n, Y_1, Y_2, Y_3, \dots, Y_m$ опорные слова (невыделенные либо выделенные в процессе обработки) и фрагменты текста; A, B – некоторые предикаты.

Последовательность опорных слов также можно представить в виде:

$$A_1(X_1, X_2), A_2(X_2, X_3), \dots, A_{n-1}(X_{n-1}, X_n)$$

$$B_1(X_1, X_2), B_2(X_2, X_3), \dots, B_{n-1}(X_{n-1}, X_n)$$

и т.д.

Представляя предикатное описание графически, получаем последовательность опорных слов в виде семантической сети рис.2.

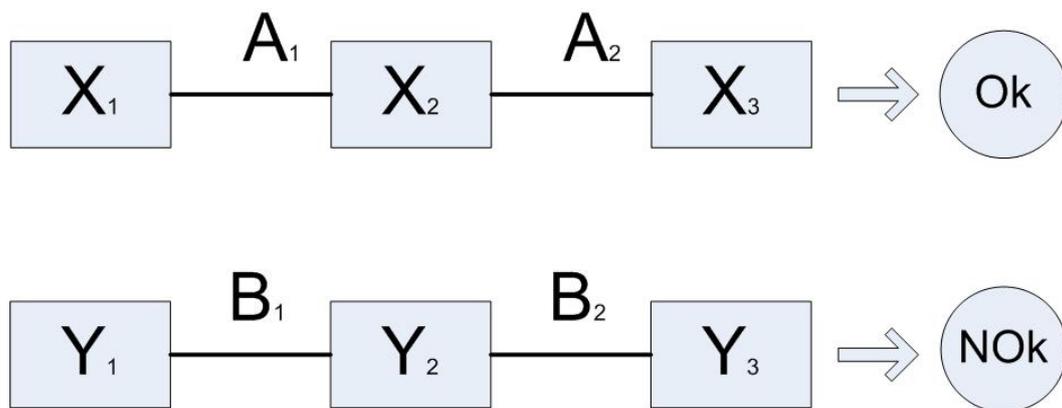


Рис. 2

Символами «Ok» и «NOK» на Рисунок 2 показано решение эксперта задачи о сортировке, т.е. решение о том, что данный документ/текст относится (или не относится) к выбранной альтернативе.

Краткие выводы

Обобщая все вышеизложенное можно сделать следующие выводы:

- 1) Представляя отношения дугами графа, мы визуализируем алгоритм рассуждения эксперта о предметной области.
- 2) Двигаясь по дугам графа мы приходим к результату – решению эксперта в задаче о сортировке.
- 3) Расширяя множество $X_1, X_2, X_3, \dots, X_n, Y_1, Y_2, Y_3, \dots, Y_m$ опорных слов с помощью специализированного словаря мы можем получить совокупность продукционных правил „ЕСЛИ-ТО”.

Благодарности

Автор благодарен доценту к.ф.н. Леандру Литову, Заведующему секцией Атомная Физика, Физический Факультет, Софийский Университет „Св. Климент Охридски”, Болгария, за предоставленный доступ к компьютерным ресурсам грид-структуры CERN.

Библиография

- [Baddeley, 2009] Baddeley, A.D., Eysenck, M., Anderson, M.C. Memory. Hove: Psychology Press. 2009.
- [Luger, 2002] George F. Luger. Artificial Intelligence. Structures and Strategies for Complex Problem Solving, 4th Edition, Addison Wesley. 2002.
- [Winograd, 1972] Understanding Natural Languages. New York: Academic Press, 1972
- [Атанасова и др., 2010] Атанасова Т., Савченко Т.Н., Головина Г.М., Баканов А.С. Интеллектуальная информационная среда обитания и субъективное восприятие качества жизни // Методы исследования психологических структур и их динамики. Труды ИП РАН. М., 2010.
- [Баканов, 2009] Баканов А.С. Особенности психологического подхода к моделированию человеко-компьютерного взаимодействия // Вестник ГУУ. 2009. №6. С. 15–18.
- [Баканова, 2007] Баканова Н. Б. Использование программно-технических комплексов для повышения эффективности контроля в системах документооборота // «Электросвязь». 2007. № 6. С. 51–53.
- [Величковский, 2006] Величковский Б.М. Когнитивная наука: Основы психологии познания. В 2 т. Т. 1. М.: Смысл, 2006.
- [Ларичев, 1987] Ларичев О.И., Петровский А.Б. Системы поддержки принятия решений. Современное состояние и перспективы развития. // Итоги науки и техники. Серия Техническая кибернетика. М. ВИНТИ, 1987. т.21, с.131-164.
- [Моросанова, Индина, 2011] Моросанова В.И., Индина Т.А. Регуляторные и личностные основы принятия решений. – СПб.; М.: Нестор-История, 2011, - 282 с.
- [Петровский, 2004] Петровский А.Б. Многокритериальное принятие решений по противоречивыми данным: подход теории мультимножеств. // Информационные технологии и вычислительные системы, 2004, №2, 56-66.
- [Ташев и др., 2013] Ташев Т., Баканова Н., Ташева Р. Исследование верхней границы пропускной способности коммутационного узла при входящем трафике типа “горячей точки”. International Journal "Information Technologies & Knowledge", Vol.7, No. 2, 2013, pp. 182-189.
- [Холодная, 2002] Холодная М. А. Когнитивные стили: О природе индивидуального ума. Учебное пособие – М.: ПЕР СЭ, 2002. -304 с.

Информация об авторах



Баканов Арсений Сергеевич – к.т.н., Институт психологии Российской академии наук, Москва, Россия; e-mail: arsb2000@pochta.ru

Основные области научных исследований: Инженерная психология, Системы принятия решений, Компьютерные технологии

Psychological Approach to Extracting Knowledge in Human-Computer Interaction with Electronic Docflow System

Arseniy Bakanov

Abstract: This article describes the method of extraction and visualization of knowledge in the process of human-computer interaction with electronic docflow system. The paper describes the experiments, the results of experiments are given. The described approach to extracting knowledge, focused on the use of docflow systems, and involves the use of algorithms and rules to identify heuristics in decision-making in solving management problems.

Keywords: Docflow system, Extracting knowledge, Visualization knowledge.