



УДК 004.8

ИЗВЛЕЧЕНИЕ ЗНАНИЙ ПОЛЬЗОВАТЕЛЯ И ВЕРИФИКАЦИЯ ЗНАНИЙ САМООРГАНИЗУЮЩИХСЯ ИНФОРМАЦИОННЫХ СИСТЕМ С ОБРАТНОЙ СВЯЗЬЮ

Диковицкий В.В.*

* Учреждение Российской академии наук

Институт информатики и математического моделирования технологических процессов

Кольского научного центра РАН,

г. Анатиты, Российская Федерация

dikovitsky@iimm.kolasc.net.ru

В работе предложен метод формализации различий знаний о предметной области между пользователями и базой знаний информационной системы, и их использование для поддержки актуальности базы знаний.

Ключевые слова: база знаний, интеллектуальная система, предметная область, семантическая сеть.

ВВЕДЕНИЕ

Актуальной на сегодняшний день является задача обработки информационными системами различий знаний о предметной области как между пользователями, как и между конкретным пользователем и базой знаний (БЗ) информационной системы (ИС). В работе [Gruber, 1995], данное различие трактуется как проблема человеко-машинного и межкоммуникантного взаимодействия. Одной из главных задач разработки онтологий является выработка единого словаря для пользователей ИС [Pinto, 1999]. Данная задача актуальна для формализации общепринятых знаний, и не может рассматриваться в рамках мультипредметных динамических предметных областей (ПрО). В реальных условиях часто даже общепринятые знания трактуются разными пользователями по-разному, и разработка единого словаря для них не только нецелесообразна, но и невозможна. Различные представления пользователей о ПрО представляется возможным использовать для получения новых и верификации имеющихся знаний, что особенно актуально для машинно-добавляемых знаний. В работе представлен метод динамического формирования БЗ на основе пользовательских знаний, позволяющий динамически в автоматическом режиме формировать пространство понятий пользователей ИС.

1. Извлечение знаний пользователя

Модель пользователя ИС – это, прежде всего его информационная модель, которая состоит из множества понятий и отношений между ними (1), основное назначение модели пользователя –

хранить область интересов конкретного пользователя.

$$UM = \langle T, L, W_T \rangle, \quad L = \langle T_i, T_j, W \rangle, \quad (1)$$

где T – множество тем для поиска, соответствующих концептам базы знаний, W_T – вес концепта в модели конкретного пользователя, L – множество отношений над концептами, W – вес отношений.

Структура модели пользователя (1) изоморфна структуре БЗ ИС, т.е. модель пользователя – это своего рода БЗ, содержащая знания пользователя, известные информационной системе. Следовательно, представляется возможным рассматривать отличные от БЗ представления пользователя о предметной области как новые знания:

$$NK = \langle T', L', W'_T \rangle, \quad (2)$$

где T' – множество тем для поиска, не соответствующих концептам базы знаний, W'_T – вес термина в модели конкретного пользователя.

Таким образом, различные представления пользователей о предметной области – это разница между информационной моделью пользователя и базой знаний ИС. Суммирование новых знаний (2) позволяет выявить синонимы концептов БЗ, новые виды в классификации понятий и т.п. Извлекать пользовательские знания представляется возможным путем обработки запросов пользователя, а так же путем учета статистики его взаимодействия с динамическими навигационными структурами (Рисунок 1).

Запрос можно представить как множество тем для поиска, задающих информационную потребность пользователя.

$$UQ = \langle T, W_T \rangle \quad (3)$$

где T - множество ключевых слов - тем для поиска, W_T - вес концепта в модели конкретного пользователя.

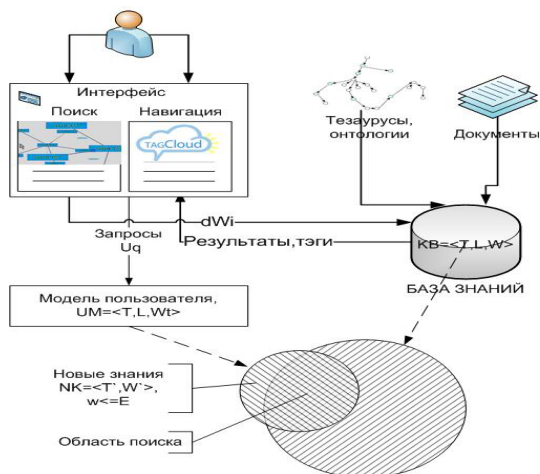


Рисунок 1 - Извлечение знаний пользователя

При пересечении множества (3) с БЗ, часть модели пользователя, непересекающаяся с БЗ, представляет новые знания(4), дополняющие БЗ.

$$UM / KB = \langle T', L', W' \rangle = NK, W' \leq \varepsilon \quad (4)$$

Вес связей при добавлении минимален и получает приращение с каждым обращением. Таким образом, связи являются предпосылками, и, в случае обнаружения ИС документа, имеющего пересечение с данной структурой (4) – вес связей преодолевает порог отображения ε (количественная характеристика, соответствующая весу отношений между понятиями БЗ, при котором они доступны для пользователей ИС). Это позволяет, например, выявить и добавить в БЗ отсутствующие синонимы понятий, описанных в документах, отсутствующие категории классификации и отображать их при следующих обращениях к БЗ. Модель пользователя хранит область интересов пользователя и позволяет определить контекста

2. Модификация базы знаний

Поддержка актуальности хранимых знаний представляется возможной путем выработки воздействий, направленных на модификацию БЗ - удаление и добавление объектов базы знаний. Удаление концепта семантической сети происходит при устремлении к нулю весов отношений - несвязные концепты подлежат удалению. Добавление новых знаний (4) к семантической сети происходит с весами, не превышающими порог отображения ε , т.е. новые знания, представленные в виде различий между представлений о ПрО в модели пользователя и БЗ, доступны только для компонентов приобретения знаний.

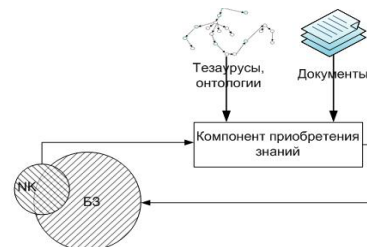


Рисунок 2 – Модификация базы знаний

Таким образом, знания NK существуют в виде предпосылок для компонента приобретения знаний, и подтверждение данных знаний является критерием добавления в БЗ.

ЗАКЛЮЧЕНИЕ

В работе предприняты попытки формализации знаний пользователя как различных его представлений о ПрО и их использование для формирования модели пользователя, основная задача которой – уменьшение неопределенности предмета поиска и верификации БЗ.

Верификация БЗ актуальна для систем, использующих автоматизированные компоненты приобретения знаний. Обработка естественного языка влечет за собой неопределенность, снизить которую представляется возможным с помощью обратной связи с пользователем. Кроме того, данный метод позволяет рассматривать различия модели пользователя и базы знаний как новые знания, что является одним из аспектов развития человеко-машинного взаимодействия.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

[Gruber, 1995] Gruber T. R. «Toward principles for the design of ontologies used for knowledge sharing». Presented at the Padua workshop on Formal Ontology, March 1993

[Pinto, 1999] Pinto HS, Gómez-Pérez A, Martins JP (1999) Some Issues on Ontology Integration. / In Proc. of IJCAI99's Workshop on Ontologies and Problem Solving Methods: Lessons Learned and Future Trends. Vol. 18, pp. 1 - 12. Stockholm, Sweden, 1999.

EXTRACTION OF USER KNOWLEDGE AND VERIFICATION OF KNOWLEDGE SELF-ORGANIZING INFORMATION SYSTEMS WITH FEEDBACK

Dikovitsky V.V. *

* Establishment of Russian Academy of Sciences Institute for Informatics and Mathematical Modelling of Technological Processes of the Kola Science Center RAS, Apatity, Russian Federation

dikovitsky@iimm.kolasc.net.ru

In this paper we propose a method of formalizing differences between knowledge of users and knowledge base of information systems, and using this differences for verification knowledge base.