



УДК 004.023

ИССЛЕДОВАНИЕ ВРЕМЕНИ ПОИСКА В ТЕМПОРАЛЬНЫХ БАЗАХ ЗНАНИЙ С ДИНАМИЧНЫМ КОНТЕНТОМ

Ю.А. Королёва^а^а Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

Адрес для переписки: julietkoroleva@gmail.com

Информация о статье

Поступила в редакцию 06.02.15, принята к печати 21.05.15

doi:10.17586/2226-1494-2015-15-4-764-766

Язык статьи – русский

Ссылка для цитирования: Королёва Ю.А. Исследование времени поиска в темпоральных базах знаний с динамичным контентом // Научно-технический вестник информационных технологий, механики и оптики. 2015. Т. 15. № 4. С. 764–766.**Аннотация**

Предложены результаты исследования эффективности поиска актуальных данных в темпоральных базах знаний, построенных в базисе состояния–события и позволяющих, наряду с быстрым доступом к актуальным состояниям, обращаться к предыстории на основе хронологии событий. Показано, что хранение данных на глубокую ретроспективу приводит к существенному увеличению времени поиска за счет роста дерева решений. Исследовано время поиска в темпоральных базах знаний в зависимости от среднего числа событий, предшествующих текущему состоянию. Полученные результаты подтверждают преимущество баз знаний в базисе состояния–события перед традиционными методами построения интеллектуальных систем.

Ключевые слова

информационный поиск, динамичная база знаний, темпоральная база знаний.

RETRIEVAL TIME RESEARCH IN TEMPORAL KNOWLEDGE BASES WITH DYNAMIC CONTENT

J. A. Koroleva^а^а ITMO University, Saint Petersburg, 197101, Russian Federation

Corresponding author: julietkoroleva@gmail.com

Article info

Received 06.02.15, accepted 21.05.15

doi:10.17586/2226-1494-2015-15-4-764-766

Article in Russian

For citation: Koroleva J.A. Retrieval time research in temporal knowledge bases with dynamic content. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2015, vol.15, no. 4, pp. 764–766.**Abstract**

Results of retrieval time research of actual data effectiveness search in temporal knowledge bases built in the basis of state of events have been proposed. This type of knowledge base gives the possibility for quick access to relevant states as well as for history based on events chronology. It is shown that data storage for deep retrospective increases significantly the search time due to the growth of the decision tree. The search time for temporal knowledge bases depending on the average number of events prior to the current state has been investigated. Experimental results confirm the advantage of knowledge bases in the basis of state of events over traditional methods for design of intelligent systems.

Keywords

information retrieval, dynamic knowledge base, temporal knowledge base.

Информационный поиск в базах знаний (БЗ) на продукционной модели обладает экспоненциальной сложностью от глубины дерева решений, вследствие чего требуется применение специальных алгоритмов для редуцирования дерева поиска, что особенно актуально для систем поддержки принятия решений, в частности, систем управления сложными техническими объектами [1, 2], систем управления перевозками, логистики и др.

Темпоральная БЗ содержит исторические (хронологические) данные, т.е. данные, относящиеся к прошлым и, возможно, к будущим периодам времени. При добавлении утверждений происходит простое добавление в конец файла новых фактов. Таким образом, с течением времени количество фактов, связанных с объектом, увеличивается. Резолюция цели при такой организации БЗ влечет за собой спуск по дереву решений, что приводит к увеличению времени поиска. Следовательно, контроль глубины ретроспективы событий для обеспечения заданных ограничений на время информационного поиска является актуальной проблемой.

Одним из путей существенного сокращения времени информационного поиска является сохранение результатов предыдущего поиска в качестве прецедентов [3–5]. Такой подход может использоваться в тех случаях, когда БЗ не изменяется, но не может применяться для динамических БЗ, поскольку прецеденты могут оказаться неактуальными.

Динамичность БЗ означает возможность появления новых, изменения или удаления существующих фактов [6, 7]. В этих случаях прецеденты как результаты предыдущих запросов к БЗ становятся неактуальными. Перечисленные случаи соответствуют модели знаний экспертных систем, в которых факты являются вторичными по отношению к правилам [5]. В работе [8] автором предложено рассматривать темпоральные БЗ как разновидность динамических БЗ, в которых предлагается различать первичные факты, которые не могут ни исчезать, ни изменяться, и вторичные факты, которые являются результатами применения продукционных правил. Первичный факт соответствует некоторому событию, а вторичный – состоянию. Таким образом:

- все многообразие фактов в базах знаний состоит из множества состояний $S = \{s\}$ и множества событий $E = \{e\}$, обуславливающих переходы из одного состояния в другое с помощью правил $R = \{r\}$;
- определение состояния объекта в произвольный момент времени требует вычисления либо всех переходов от начального состояния до текущего путем обработки множества событий об объекте, предшествующих этому моменту, либо всех переходов от запомненного состояния до текущего момента;
- запомненное текущее состояние объекта s_t является текущим до тех пор, пока не произойдет событие e_{t+1} , изменяющее состояние данного объекта.

Ниже приведена упрощенная схема БЗ в пространстве «состояния–события», описанная в [7] и допускающая создание универсальных правил для обновления состояний при появлении новых событий. Определим состояния и события Prolog-предикатами status и event соответственно:

status(<subject>, <predicate>, <object>, <date>). event(<subject>, <event>, <date>).

Аргумент date в предикате status соответствует date, начиная с которой данное состояние является истинным. Для неизменных состояний эта дата обязательно должна быть раньше любого события в базе знаний.

Определим также допустимые переходы из одного состояния в другое: jmp(<class>, <event>, <oldstate>, <newstate>).

Предикат jmp определяет класс субъектов и события, к которым применимы переходы из состояния oldstate в состояние newstate. Например, для класса person и семейного положения (maritalStatus) допустимыми являются следующие переходы:

```
jmp(person, birth, maritalStatus, _, single).
jmp(person, divorce, maritalStatus, married, divorced).
jmp(person, marriage, maritalStatus, single, married).
jmp(person, marriage, maritalStatus, divorced, married).
jmp(person, marriage, maritalStatus, widowed, married).
jmp(person, spouseDeath, maritalStatus, married, widowed).
```

Рассмотрим две базы знаний – темпоральную БЗ и упрощенную БЗ, в которой сохраняется только текущее состояние без возможности ретроспективного поиска (БЗ текущих состояний).

В БЗ текущих состояний хранятся только последние актуальные состояния, которые обновляются при наступлении событий в соответствии с допустимым множеством переходов.

Обе БЗ построены с помощью генератора случайных чисел, количество объектов в них варьируется от 10 000 до 70 000, каждому объекту сопоставлено 5 атрибутов, задающих их признаки. Количество объектов, соответствующих цели поиска (релевантных поиску), составляет 7/8 от общего количества объектов.

Поисковый запрос задан правилом вида
Утверждение1, Утверждение2, ..., УтверждениеN.

В процессе информационного поиска каждое утверждение запроса может унифицироваться либо с фактом, либо с головой другого правила. Таким образом, обеспечивается вложенность правила. Путем варьирования количества антецедентов может быть изменен коэффициент ветвления дерева решений. В поставленном эксперименте коэффициент ветвления дерева решений (k) варьировался от 2 до 4. Глубина поиска зависит от количества хронологических данных, которые внесены в БЗ.

В результате вычислительного эксперимента получены зависимости времени поиска от числа объектов в БЗ, к которым привязываются события и состояния. Число событий, относящихся к одному объекту, определяет коэффициент ветвления k . Для БЗ текущих состояний коэффициент ветвления всегда равен 1. На рисунке отображено отношение времени T поиска в БЗ с различной глубиной ретроспективы к времени T_0 поиска в БЗ текущих состояний. Представленные на рисунке данные демонстрируют, что наблюдаемое увеличение времени поиска в большей степени характерно для небольших баз знаний. С ростом числа объектов это отношение уменьшается, но, тем не менее, остается достаточно большим.

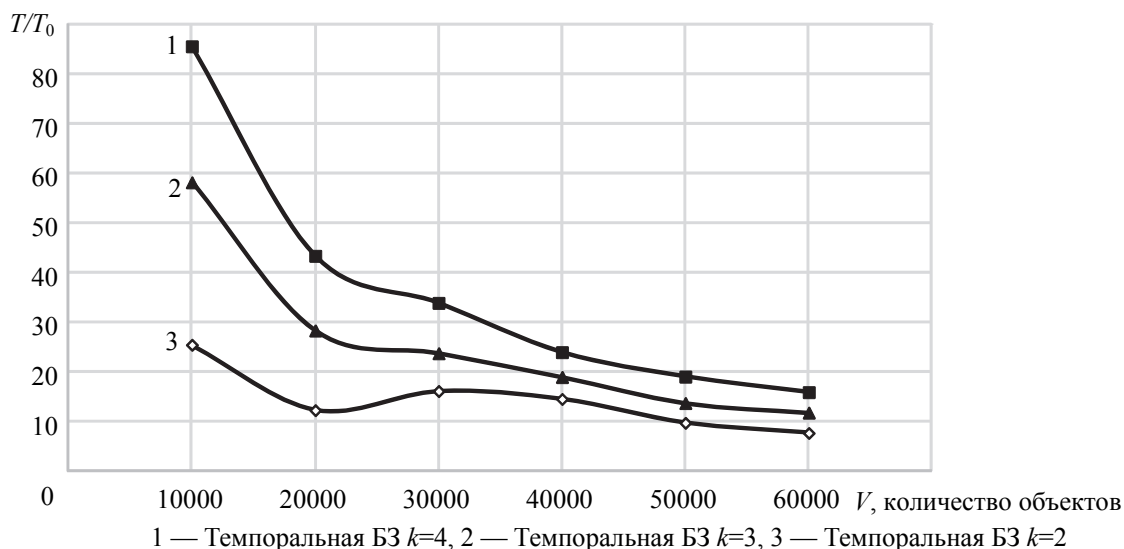


Рисунок. Увеличение времени поиска в зависимости от глубины ретроспективы

В результате проведенного исследования установлено, что БЗ, хранящая в явном виде последние актуальные состояния, обеспечивает абсолютный выигрыш во времени поиска по сравнению с темпоральными БЗ, хранящими историю состояний. Увеличение глубины ретроспективы приводит к существенному росту времени поиска, особенно в небольших БЗ. Начиная с четырех ретроспективных событий это время приблизительно в 15–50 раз выше, чем время поиска в БЗ текущих состояний. Таким образом, полученные результаты демонстрируют целесообразность построения БЗ в базе состояния–события, кроме случаев, когда ретроспективный поиск является обязательным.

1. Еремеев А.П., Куриленко И.Е. Компонента временных рассуждений для интеллектуальных систем поддержки принятия решений реального времени // Искусственный интеллект и принятие решений. 2009. № 1. С. 31–45.
2. Eremeev A.P., Kurilenko I.E. Temporal reasoning component for real-time intelligent decision-support systems // Journal Scientific and Technical Information Processing. 2011. V. 38. N 5. P. 332–343. doi: 10.3103/S0147688211050030
3. Катериненко Р.С., Бессмертный И.А. Метод ускорения логического вывода в продукционной модели знаний // Программирование. 2011. Т. 37. № 3. С. 76–80.
4. Katerinenko R.S., Bessmertnyi I.A. A method for acceleration of logical inference in the production knowledge model // Programming and Computer Software. 2011. V. 37. N 4. P. 197–199. doi: 10.1134/S0361768811040049
5. Bessmertny I. An intellectual agent in training systems // Proc. IMSCI 2007 - International Multi-Conference on Society, Cybernetics and Informatics. Orlando, USA, 2007. P. 86–89.
6. Xu Z., Liu Y., Mei L., Hu C., Chen L. Generating temporal semantic context of concepts using web search engines // Journal of Network and Computer Applications. 2014. V. 43. P. 42–55. doi: 10.1016/j.jnca.2014.04.002
7. Verduijn M., Dagliati A., Sacchi L., Peek N., Bellazzi R., de Jonge E., de Mol B. Comparison of two temporal abstraction procedures: a case study in prediction from monitoring data // AMIA ... Annual Symposium proceedings. 2005. P. 749–753.
8. Бессмертный И.А., Королёва Ю.А. Способ организации баз знаний, ориентированный на создание и использование прецедентов // Труды конгресса по интеллектуальным системам и информационным технологиям IS&IT'14. Дивноморское, 2014. С. 245–250.

Королёва Юлия Александровна – аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, julietkoroleva@gmail.com

Julia A. Koroleva – postgraduate, ITMO University, Saint Petersburg, 197101, Russian Federation, julietkoroleva@gmail.com