

УДК 004.827:902.6]:069.51

Г.Ф. Крисилова, науч. сотр.,
Одес. археологич. музей,
О.В. Липская, магистр,
Одес. нац. политехн. ун-т.

РЕШЕНИЕ ЗАДАЧ ХРАНЕНИЯ И ПОИСКА ВРЕМЕННЫХ ХАРАКТЕРИСТИК ЭКСПОНАТОВ АРХЕОЛОГИЧЕСКО- ГО МУЗЕЯ

Г.Ф. Крісілова, О.В. Липська. Задачі збереження та пошуку часових характеристик археологічного музею. Розглянуто задачі збереження та пошуку часових характеристик археологічного музею. Розроблено метод збереження часових характеристик, що базується на понятті лінгвістичної змінної. Розроблено показник, який допомагає формувати релевантну вибірку експонатів за часовими характеристиками, обчислюючи міру відповідності часової характеристики експоната по відношенню до еталона.

G.F. Krissilova, O.V. Lipska. Time characteristics' storage and search problems of archeologic museum exhibits. Exhibits storage and search problems of the archeologic museum using time characteristics are considered. The storage method of time characteristics based on the concept of a linguistic variable is developed. The parameter which calculated the conformity degree of time characteristics in relation to reference is developed. This parameter helps to form a relevant exhibits sample using time characteristics.

При хранении, учете и обработке экспонатов археологического музея большое значение имеет анализ временных характеристик — временных интервалов, которыми датируется каждый экспонат. Временные характеристики имеют нечеткий характер в силу специфики предметной области археологии. В большинстве случаев невозможно точно определить год. Можно говорить лишь о некотором временном интервале, определяющемся такими словесными формулировками, как “век”, “половина века”, “треть века” и т.д. Такие формулировки неточны и имеют разную степень детализации, что существенно затрудняет анализ временных характеристик объекта. Часто лишь приблизительно известно, когда произошло интересующее событие. В общем случае, временную неопределенность можно характеризовать как “неизвестно, когда произошло”, или более конкретно — “неизвестно, когда точно произошло”.

Существует несколько методов, использующихся при работе с неопределенными событиями.

Первый применяется для решения задач, в которых присутствует временная неопределенность, и является вероятностным [1]. Он предназначен для работы только со специфическими показателями, определяемыми в терминах экономической теории, таких как: “бизнес-день”, “бизнес-неделя”, “квартал”. Кроме того, существенным недостатком метода является необходимость в статистических исследованиях, на основании которых должна определяться вероятностная функция возникновения события в текущий период времени. При отсутствии таких исследований адекватность результатов сомнительна.

Второй метод основан на грануляции [2]. Основополагающим в данном методе является предположение, что интервалы времени четкие и имеют непересекающиеся границы, в то время как специфика задачи анализа археологических экспонатов предполагает нечеткую информационную грануляцию, возникающую вследствие высокой неопределенности и субъективности задания временных лингвистических переменных. Метод неэффективен при работе с временными характеристиками экспонатов.

До сих пор хранение временных характеристик экспоната сводилось к указанию в его учетной карточке строковой информации о датировке. Такой способ ведет к увеличению объемов хранимых данных и делает затруднительным анализ временных характеристик, которые несут смысловую нагрузку, связанную с их числовыми значениями, и обладают высокой неопределенностью. Следовательно, поиск экспонатов по датировке существенно затрудняется.

Задача хранения временных характеристик экспоната заключается в нахождении такого способа хранения и представления нечетких датировок, который бы позволил обеспечить формализованный ввод и снизить объем хранения данных, а также повысить их информативность.

Разработана методика, позволяющая формировать релевантную выборку экспонатов с нечеткими датировками по заданной нечеткой датировке, являющейся эталонной. В исследовании используется метод, предполагающий нечеткое гранулирование.

Нечеткая информационная грануляция лежит в основе базового понятия нечеткой логики – понятия лингвистической переменной [3]. Нечеткая датировка представляется в виде лингвистической переменной, а затем производится ее дальнейший анализ.

Лингвистическая переменная характеризуется набором

$$X, T(X), U, G, M,$$

где X — название переменной;

$T(X)$ — терм-множество переменной X , т.е. множество названий ее лингвистических значений, причем каждое из этих значений является нечеткой переменной X со значениями из универсального множества U с базовой переменной u ;

G — синтаксическое правило, порождающее названия значений переменных X ;

M — семантическое правило, которое ставит в соответствие каждой нечеткой переменной X ее смысл $M(X)$.

Любая датировка рассматривается как лингвистическая переменная, определенная на интервале $0 \dots 100$, соответствующем длине века. Значениями лингвистической переменной датировки являются нечеткие переменные с собственной областью определения и функцией принадлежности. Область определения такой переменной представляет собой временной интервал с точностью до года.

Временной интервал — это непрерывный отрезок времени, определяющийся началом и концом.

Введем обозначения: t_n — год начала интервала; t_k — год конца интервала; $\mu(t_i)$ — значение функции принадлежности в точке $t_i \in t_n \dots t_k$.

Сформулируем принцип непрерывности, который должен соблюдаться при задании датировки: внутри интервала $t_n \dots t_k$ не существует точек, степень принадлежности которых данной датировке равна 0.

$$\forall t_i | t_i \in t_n \dots t_k \Rightarrow \mu(t_i) > 0.$$

Любая датировка, представляющая собой лингвистическую переменную, строится согласно принципу непрерывности. Значения датировки задаются трапецевидной функцией принадлежности, наиболее часто применяемой в теории нечетких множеств [3], с вектором параметров нечеткой переменной a, b, c, d , удовлетворяющих условию $a \leq b \leq c \leq d$.

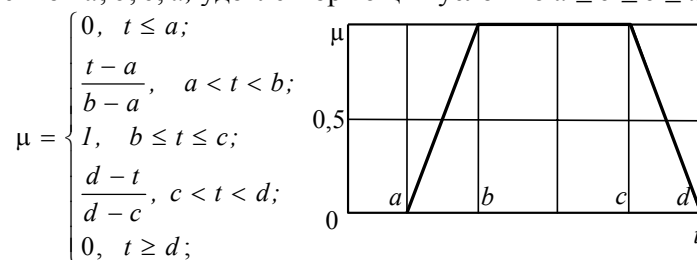


Рис. 1. Трапецевидная функция принадлежности

Первая задача при работе с временными характеристиками экспоната — хранение датировки.

Временные характеристики хранятся с различной степенью подробности, зависящей от возможности точной датировки каждого экспоната. Наиболее точным является указание года. Однако в большинстве случаев для задания датировок применяются такие понятия, как “четверть века”, “треть века”, “часть века”, “половина века”, “век”. Эти понятия имеют различную

степень подробности, кроме того, определены в различных шкалах, связанных между собой и со шкалой времени. Предлагается хранить временные характеристики экспоната в виде значений лингвистических переменных. Формирование лингвистических переменных датировки будет представлять собой переход от абсолютной наиболее точной шкалы “годы”, с шагом в один год, к относительным шкалам, выбор которых зависит от временных характеристик.

Используемые лингвистические переменные датировки и их значения можно представить в виде нечетких множеств:

- “век”; поскольку рассматривается временной домен $0 \dots 100$, этой лингвистической переменной соответствует единственное значение нечеткого множества “век”;
- “половина века”=(“первая”, “вторая”);
- “треть века”=(“первая”; “вторая”; “последняя”);
- “четверть века”=(“первая”; “вторая”; “третья”; “последняя”);
- “часть века”=(“начало века”; “середина века”; “конец века”).

Таким образом, в справочной таблице для предметной области будут храниться все возможные значения пяти лингвистических переменных датировки с уникальным вектором параметров, соответствующим конкретной лингвистической переменной.

Формы задания временных характеристик экспонатов различны. Например, лингвистическая переменная X =“треть века”, $T(X)$ =(“первая треть”, “вторая треть”, “последняя треть”), $U=0 \dots 100$. Тогда датировка экспоната “вторая треть 14-го века” будет задана в виде значения лингвистической переменной “треть века”=“вторая треть” и самого числового значения века — 14. Зачастую датировка задается в виде интервала, ограниченного снизу и сверху перечисленными терминами, называемыми начальным и конечным значениями датировки, соответственно. Например, при датировке “первая половина III века ... вторая половина V века” начальное значение датировки — “первая половина III века”, конечное — “вторая половина V века”. Кроме того, каждому начальному и конечному значениям датировки соответствует булево значение (до/после — 0/1) нашей эры.

Для хранения временных характеристик экспоната необходимо хранить:

- вид лингвистической переменной для его начальной и конечной датировки;
- начальное и конечное числовое значение веков;
- булево значение (0/1) нашей эры для начального и конечного значений датировки.

До применения предложенной методики информация о датировке экспоната хранилась в виде строковой информации, как строка из 50-ти символов. Использование метода хранения датировок в виде лингвистических переменных дает уменьшение объема хранимой информации приблизительно в 6 раз.

Второй задачей при работе с временными характеристиками экспонатов является поиск экспоната по эталонной датировке.

На первом этапе выполнения поиска определяется некоторый необходимый — эталонный — временной интервал. На втором — поиск датировок, которые ему соответствуют. Эталонная датировка, очевидно, должна задаваться в тех же терминах, что и датировка экспоната, т.е. в виде значения или интервала значений указанных лингвистических переменных. Тогда, по аналогии с датировкой экспонатов, наиболее точным значением эталонной будет “год”, наименее — “век”.

Для поиска экспоната по датировке необходимо проанализировать ее соответствие эталонной. Целесообразно вычислять степень этого соответствия S в диапазоне $0 \dots 100$ %.

Анализ проводится в два этапа:

- по заданному требованию определяется нечеткая лингвистическая переменная, соответствующая эталонной датировке;
- выполнение логической операции “Пересечение”, т.е. получение области пересечения (ОП), над нечеткими переменными эталонной (ЭД) и текущей анализируемой (ТД) датировками и вычисление степени соответствия временной характеристики экспоната требованию.

Возможны варианты расположения функций принадлежности ЭД и ТД экспоната на области определения (рис. 2).

ЭД и ТД не пересекаются, т.е. временная характеристика текущего экспоната полностью не соответствует требованию. Область пересечения (ОП) не существует, $S=0$.

ЭД полностью поглощает ТД. $S=100\%$. Такое происходит в двух ситуациях:

— ЭД обладает высокой неопределенностью, например, задана на большом временном интервале;

— ТД определено более точно по сравнению с ЭД, например, ЭД определена как III ... VI вв. н.э., а ТД — V вв. н.э., тогда ОП=ТД.

ТД полностью поглощает ЭД, $0 < S < 100\%$, если:

— ТД определено с высокой степенью неопределенности — на большом временном интервале и с высокой степенью нечеткости;

— очень точно задана ЭД, например, выбран точный год, поскольку неопределенность ТД выше, чем неопределенность ЭД; так, если ТД= III...VI вв. н.э., а ЭД= V в. н.э., то ТД будет соответствовать ЭД только на области пересечения — V в. н.э. При этом, чем больше ОП совпадает с ТД, тем больше степень соответствия ТД ЭД.

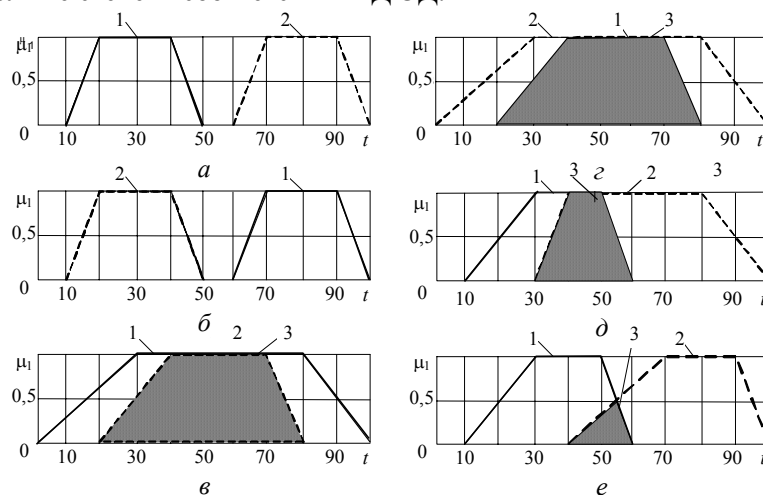


Рис. 2. Возможные варианты расположения ЭД (1), ТД (2) и ОП (3): а, б — не пересекаются, $S=0$; в — ЭД полностью поглощает ТД, $S=100\%$; г — ТД полностью поглощает ЭД, $0 < S < 100\%$; д, е — некоторая степень соответствия в ОП, $0 < S < 100\%$

Частным случаем такого соответствия является полное совпадение ТД с ЭД, т.е. ОП=ТД=ЭД, $S=100\%$.

Соответствие ТД ЭД с некоторой степенью, определяемой степенью неопределенности ОП и степенью соответствия ОП текущей датировке. $0 < S < 100\%$.

Чем больше и точнее ОП, т.е. больше область, где $\mu=1$, и при этом, чем ближе область определения ОП к области определения ТД, тем больше соответствует ТД ЭД.

Введем определение мощности множества.

Мощностью множества A M_A называется кардинальное число значений из этого множества.

Степень соответствия S ОП к ТД в процентном соотношении целесообразно определять через их мощности по формуле

$$S = \frac{M_{\text{ОП}}}{M_{\text{ТД}}} 100\%$$

При этом справедливо неравенство $0 \leq M_{\text{ОП}} \leq M_{\text{ТД}}$.

$S=100\%$, если $M_{\text{ТД}}=M_{\text{ОП}}$.

$S=0$, если $M_{\text{ОП}}=0$.

В общем случае целесообразно задавать не только вид эталонной датировки, но еще и некоторое предельное значение степени соответствия ТД ЭД S^* , чтобы в результирующую вы-

борку включались только те экспонаты, датировка которых соответствует эталону со степенью соответствия $S \geq S^*$. Тогда результирующая выборка будет релевантной.

Предложенная методика на основе механизмов нечеткой логики позволяет осуществлять поиск экспонатов; разработанная количественная характеристика позволяет оценивать соответствие датировки экспоната эталонной.

Предложенный метод формализации представления датировки экспонатов в виде лингвистических переменных позволяет уменьшить объем хранимой информации о датировках экспоната.

Для решения задачи поиска экспонатов по датировке было предложено формализовать требования, на соответствие которым осуществляется поиск, в виде эталонной лингвистической переменной и затем анализировать область пересечения эталонной лингвистической переменной и значения лингвистической переменной датировки экспоната. Результатом такого анализа является показатель степени соответствия некоторой датировки экспоната эталону, основанный на понятии мощности нечеткого множества. Такой показатель дает возможность формировать релевантную выборку экспонатов по их нечетким датировкам, получаемую при сравнении с эталонной датировкой. При задании нижней границы такого показателя становится возможным включение в результирующую выборку экспонатов, временные характеристики которых имеют низкую степень соответствия эталону.

Литература

1. Сидоров А.А. Хронологическая модель данных для промышленных информационных систем /Сидоров А.А., Маслов Д.В. //“Исследовано в России”: электронный журнал. — 2004. — <http://zhurnal.ape.relarn.ru/articles/2004/090.pdf> — 20.02.2007
2. Bettini C. A Glossary of Time Granularity Concepts / Bettini C., Dyreson C.E., Evans W.S., Snodgrass R.T. — //”Temporal Databases: Research and Practice” — 1998. — 18.02.2007
3. L. A. Zadeh Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. — //Fuzzy Sets Syst. — Vol. 90, No. 2. — 1997 — 09.03.2007

Поступила в редакцию 27 марта 2007 г.