

УДК 004.048

## СРЕДСТВА АНАЛИЗА ФОРМАЛЬНЫХ ПОНЯТИЙ ПРИ ОПИСАНИИ ОБЪЕКТНО-ПРИЗНАКОВЫХ ДАННЫХ

Кайбасова Динара Женисбековна, *dindgin@mail.ru*

Докторант факультета информационных технологий ЕНУ им. Л.Н. Гумилева,  
Астана, Казахстан

Анализ формальных понятий является методом анализа данных, с помощью которого могут быть визуализированы данные, представленные в виде таблицы «объект - свойство», «объект - атрибут». Это достигается путем построения решёточной модели особого вида, позволяющие сохранять объектно-признаковое описание сходство группы объектов. Система "Concept Explorer" позволяет реализовать метод формального концептуального анализа данных [3].

Формальный концептуальный анализ впервые был предложен R. Wille в 1981 году [1] и нашел применение в различных сферах деятельности, таких как психология, социология, медицина, биология, лингвистика, компьютерные науки, математика, электроника, экономика и многие другие.

На сегодняшний день формальный концептуальный анализ является одним из эффективных методов анализа, поскольку он позволяет идентифицировать концептуальные структуры данных множеств посредством концептуальных правил решетки и импликации. В частности, определенный набор правил воздействия как правильные последствия, может представлять собой минимальный набор условий для достижения конкретной цели. Этот метод применяется для задач поиска и анализа информации, выявления структур и зависимостей в данных, а также для формирования гипотез [1,2]. Эти понятия строятся формально – как подмножества слов, связанных друг с другом отношением принадлежности «объект - атрибут», поэтому данный метод получил название анализ формальных понятий.

Основная математическая идея [2] анализа формальных понятий — возможность построения полной решётки по любому бинарному отношению и формализация описания понятия в виде пары (объём, содержание). Существует два способа извлечения и представления знания из анализа формальных понятий: концептуальная решетка и импликация правила. Решетки понятий применяются в направлении современного анализа текстовых данных. Решетка понятий – это модель, отражающая структуру понятий, присущих тексту [3]. В этой работе, мы предложили подход, основанный на анализе формальных понятий, для определения отношений между результатами обучения и дисциплинами посредством надлежащего последствия.

Для установления связи между объектами и их свойствами служит формальный контекст. Простой формальный контекст это тройка множество объектов, множество свойств, связь между объектами и свойствами.

Согласно [2], формальным контекстом называется тройка вида

$$K=(G, M, I),$$

где  $G$  и  $M$  – множества, а  $I$  – отношение на множестве  $G \times M$ .  $G$  представляет множество объектов,  $M$  – множество свойств, а  $gIm$  означает, что объект  $g$  обладает свойством  $m$ . Элементы  $G$  называются (формальными) объектами, а элементы  $M$  (формальными) признаками [3]. Эти множества частично упорядочены некоторыми отношениями, которых обозначим  $\subseteq$  и  $\in$ . Отношение  $I \subseteq G \times M$  говорит о том, какие объекты какими признаками обладают. Связи между объектами и атрибутами определяются следующим образом. Для подмножеств  $A \subseteq G$  и  $B \subseteq M$  объектов и атрибутов задаются отображения (функции)  $A': A \rightarrow B$  и  $B': B \rightarrow A$  определены операторы Галуа:

$$A' = \{m \in M \mid \forall g \in A (g I m)\},$$

$$B' = \{g \in G \mid \forall m \in B (g I m)\}.$$

Пара множеств  $(A, B)$  таких, что  $A' = B$ ,  $B' = A$  называют формальным понятием контекста  $K$ .

Оператор " (двукратное применение оператора ') является оператором замыкания: он идемпотентен ( $A''' = A''$ ), монотонен ( $A \subseteq B$  влечет  $A'' \subseteq B''$ ) и экстенсивен ( $A \subseteq A''$ ). Множество объектов  $A \subseteq G$ , такое, что  $A'' = A$ , называется замкнутым. Аналогично для замкнутых множеств признаков — подмножеств множества  $M$ . Пара множеств  $(A, B)$ , таких, что

$$A \subseteq G, B \subseteq M, A' = B \text{ и } B' = A,$$

называется формальным понятием контекста  $K$ . Множества  $A$  и  $B$  замкнуты и называются объемом и содержанием формального понятия  $(A, B)$  соответственно. Для множества объектов  $A$  множество их общих признаков  $A'$  служит описанием сходства объектов из множества  $A$ , а замкнутое множество  $A''$  является кластером сходных объектов (с множеством общих признаков  $A'$ ). Отношение "быть более общим понятием" задается следующим образом:  $(A, B) \geq (C, D)$  тогда и только тогда, когда  $A \supseteq C$  [3].

Понятия формального контекста  $K = (G, M, I)$ , упорядоченные по вложению объемов образуют решетку  $\mathbf{B}(G, M, I)$ , называемую решеткой понятий.

Пример, рассмотрим фрагмент модели результатов обучения дисциплин формирующих профессиональные компетенции по специальности «Информационные системы». Формальный контекст имеет представление в виде матрицы инцидентности отношения  $I$ , в которой ненулевые элементы обозначают факт принадлежности атрибута (компетенции)  $m \in M$  объекту (дисциплины)  $g \in G$ .

$G = \{\text{«Алгоритмы, структуры данных и программирование»}, \text{«Объектно-ориентированное программирование»}, \text{«Базы данных в ИС»}, \text{«Основы ИС»}, \text{«Проектирование ИС»}, \text{«Верификация, стандартизация и сертификация ПО»}, \text{«Управление разработкой программного обеспечения»}, \text{«Разработка распределенных приложений»}\}$

$M = \{\text{«Обеспечивает условия жизненного цикла информационных систем»}, \text{«Осуществляет реализацию задач на языках программирования, настройку средств разработки, проводит просмотры программного кода ПО»}\}$

«Проектирует типовые БД, разрабатывает и оптимизирует сложные SQL запросы», «Разрабатывает процедуры миграции и преобразования (конвертации) данных»}

Формальный контекст удобно предоставлять в виде объектно-признаковой таблицы. При этом запись *gIm* читается как «объект *g* обладает признаком *m*». Ряд свойств концептуальной решетки может быть определен без ее построения, с помощью только формального контекста, на основании отношения "стрелки". В системе имеется возможность визуализировать отношение "стрелки", что также можно заметить на рисунке 1.

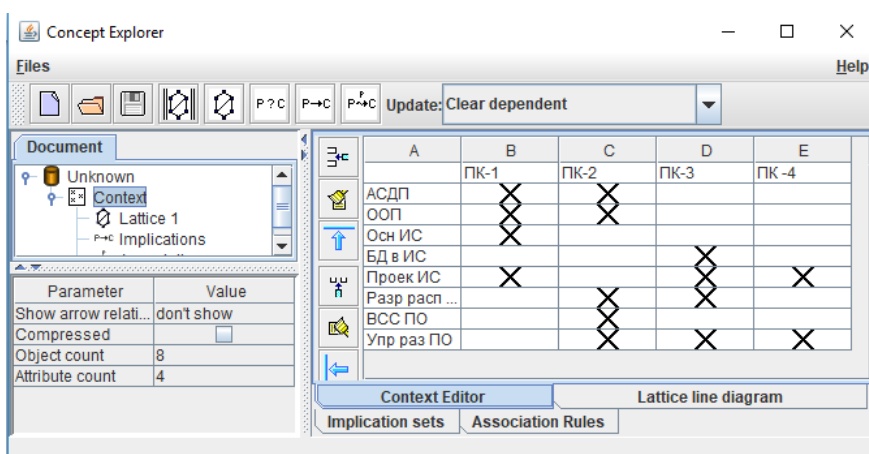
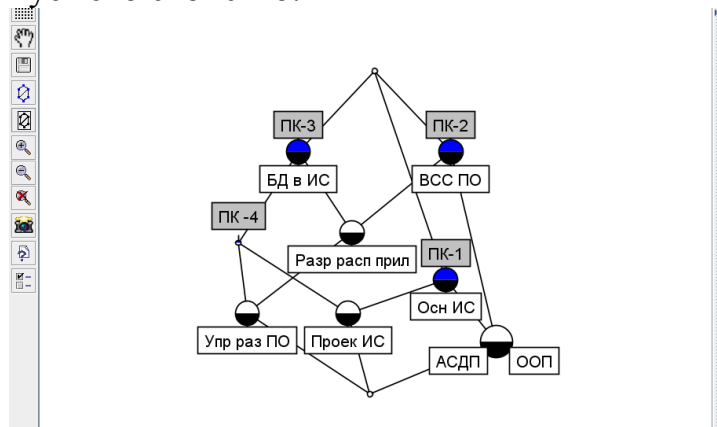


Рисунок 1 – Окно редактор Concept Explorer

Решетки понятий можно изображать с помощью обычных диаграмм Хассе. Однако, при пометке каждого понятия его объемом и содержанием изображение не будет наглядным. Поэтому используется сокращенная пометка, в которой каждый объект и каждый атрибут изображаются на диаграмме всего один раз. Имя объекта приписывается пересечению всех понятий, в объемах которых содержится этот объект, а имя свойства приписывается объединению всех понятий, содержания которых включают это свойство. Таким образом, имя объекта приписывается наименьшему из понятий, в которых встречается данный объект, а имя свойства приписывается наибольшему из понятий, в которых присутствует это свойство.



## Рисунок 2 – Линейная диаграмма концептуальной решетки

На рисунке 2 изображено окно для работы с линейными диаграммами концептуальных решеток. В нем изображена концептуальная решетка, полученная на основании формального контекста, изображенного в на рисунке 1. Так как концептуальная решетка может быть достаточно сложной, то для ее визуализации применяется эвристический алгоритм, минимизирующий количество пересечений между ребрами решетки.

В редакторе контекста Concept Explorer, кроме сохранения и редактирования формальных контекстов, также можно выполнять ряд преобразований над контекстом - устранение повторяющихся объектов с одинаковыми свойствами, и повторяющихся атрибутов, получение по контексту редуцированного контекста — контекста, в котором содержатся только объекты и атрибуты, которые не могут быть получены как комбинация других объектов или атрибутов (т.е. остаются только независимые объекты и атрибуты). Кроме того, Concept Explorer предоставляет следующие инструменты: редактирование контекста, построение концептуальных решеток из формального контекста, нахождение базы для последствий, которые являются истинными в формальном контексте, нахождение базы правил ассоциации, которые являются истинными в формальном контексте и выполнение исследования атрибутов [4].

Из вышеизложенного можно сделать вывод, что в системе, реализующей метод формального концептуального анализа, должны быть реализованы функции создания и редактирования контекстов, построения множества всех понятий и линейных диаграмм концептуальных решеток.

### **Список использованных источников**

1. Wille R. Restructuring Lattice Theory: an approach based on hierarchies of concept./ Ordered sets / editor I. Rival.- Reidel, Dordrecht-Boston , 1982.- p. 445-470.
2. Ganter, B., Stumme, G., and Wille, R. (2005). Formal concept analysis: foundations and applications, volume 3626. Springer Science & Business Media.
3. Евтушенко С.А., Система анализа данных «Concept Explorer»// Труды 7-й Национальной конференции по искусственному интеллекту. – 2000. – С.127-134.
4. Электронный ресурс: <http://conexp.sourceforge.net/download.html>