

СРАВНЕНИЕ СВОЙСТВ НОМИНАЛЬНОГО ТИПА ОБЪЕКТОВ РАЗЛИЧНЫХ ПРЕДМЕТНЫХ ПОДОБЛАСТЕЙ В РЕЛЯЦИОННЫХ БАЗАХ ДАННЫХ

М.Г. Глава

Одесский национальный политехнический университет,
пр. Шевченко, 1, Одесса, 65044, Украина; e-mail: glavamaria@mail.ru

Рассматривается проблема объединения моделей предметной области (ПрО). Предлагается сопоставлять объекты ПрО на основе значений свойств экземпляров этих объектов. Методы сопоставления свойств различаются в зависимости от типа шкал, в которых измеряются их значения. Предлагается построить онтологическую модель для свойств номинального типа и обработать ее концепты с помощью построения таблицы сопряженности признаков и метода анализа соответствий.

Ключевые слова: модель предметной области, предметная подобласть, свойства номинального типа, таблица сопряженности признаков, анализ соответствий.

Введение

На сегодняшний день работа любой организации любой сферы деятельности не возможна без использования информационных технологий. Огромный поток данных хранится и обрабатывается с помощью баз и хранилищ данных, что существенно упрощает управление и контроль деятельностью. Учитывая экономическое положение страны и анализируя рынок, подверженный реорганизации предприятий, можно сделать вывод о существовании проблемы объединения информационных систем.

Другим аспектом, подтверждающим актуальность данной проблемы, можно назвать реформы, проводимые в Украине, направленные на информатизацию учета имущественных прав, налоговой реформы и т.д. влекущие за собой создание единых информационных хранилищ страны, объединяя базы данных регионов.

Поскольку построение любой информационной системы начинается с описания предметной области (ПрО) и построения ее модели, решение поставленной проблемы сводится к объединению моделей ПрО рассматриваемых информационных систем. Результатом объединения моделей ПрО будет модель ПрО более высокого $(j+1)$ -го порядка, а модели более низкого порядка j будем называть моделями предметных подобластей (ПрПО). Для получения модели $(j+1)$ -го порядка, адекватной исследуемой ПрО [1], необходимо определить объекты, находящиеся на пересечении объединяемых ПрПО, т.е. определить подобные объекты, описываемые подобными свойствами. Предложенный подход позволит избежать избыточности и несогласованности данных ПрО $(j+1)$ -го порядка, а также потери данных, наработанных за период существования объединяемых информационных систем, за счет выделения/объединения подобных объектов/свойств объектов и дополнения модели ПрО отличающимися объектами ПрПО.

В работе [2] предложена технология поиска проекций одних и тех же ПрПО (объектов ПрПО), в которой предлагается сопоставлять объекты на основе значений свойств экземпляров этих объектов. Алгоритмы сопоставления различаются в

зависимости от типа данных конкретного свойства. В данной работе предлагается алгоритм сопоставления свойств номинального типа.

Согласно предложенной технологии объекты потенциально подобных ПрПО необходимо подготовить к сопоставлению: выделить существенные свойства, основываясь на количестве информации каждого свойства и оценке экспертов; проранжировать по значимости объекты каждой сравниваемой ПрПО, основываясь на количестве и степени важности связей определенного объекта с другими в этой же ПрПО, и количестве значимых свойств, измеряемых определенной шкалой (порядковой, номинальной, числовой); отсортировать кортежи по значениям порядковых и номинальных свойств, соблюдая полученный ранее ранг свойств.

Алгоритм сопоставления свойств номинального типа

Алгоритм сопоставления свойств номинального типа выполняется после анализа порядковых свойств. Можно предположить, что предыдущие шаги сблизили ранг потенциально подобных объектов и их свойств. А в силу работы алгоритма сопоставления свойств порядкового типа, приблизилось соответствие кортежей сравниваемых объектов, соответственно, значения свойств номинального типа сопоставляются для выровненных мощностей множеств кортежей. Под рангом понимается место объекта/свойства в последовательности рассматриваемых объектов/свойств, определяемое при помощи порядковой шкалы.

Поскольку информацию об одном и том же объекте или действии можно представить разными номинальными значениями, посимвольное сравнение отклоняется. Сравнение семантических значений экземпляров свойств для решения представленной проблемы является нетривиальной задачей и практически не выполнимой в силу своей сложности и объема работы. А также семантический анализ исключает универсальность технологии сопоставления ПрПО, поскольку потребует огромных усилий и затрат времени экспертов в конкретной ПрО/ПрПО.

Для сопоставления свойств номинального типа предлагается построить модель онтологии, которая характеризует любые номинальные свойства.

На сегодняшний день термин «онтология» не имеет устоявшегося определения, но в информационных технологиях чаще всего используется определение, сформулированное Т. Р. Грубером: «Онтология – это спецификация концептуализации» [4]. Под концептуализацией понимается представление понятий, которые классифицируют объекты ПрО и связи между ними.

Выделяют следующие типы онтологий [5–7]: метаонтология (онтология верхнего уровня) – оперирует общими концептами и отношениями, которые не зависят от конкретной ПрО; предметная онтология – содержит понятия, описывающие конкретную ПрО и отношения между ними; онтология задач – содержит функции, с помощью которых производится преобразование входных данных в выходные.

Для решения поставленной задачи будем использовать метаонтологию, поскольку необходимо создать технологию сопоставления ПрПО, независящую от самих ПрПО.

Процесс создания метаописаний также называют аннотированием, которое может происходить как с помощью человека, так и с помощью специальных алгоритмов, реализованных программно [8].

В метаописаниях выделяют три типа: системные (служебные) метаданные – предназначены для функционирования информационных систем и систем управления знаниями; структурные метаданные – содержат справочную информацию об объектах, т.е. описания, использующиеся при идентификации и категоризации объектов в тех или иных целях; семантические метаописания – включают концептуальное (аннотированное) изложение содержания и смысла информации об объекте.

Остановимся на структурных метаданных поскольку данный подтип позволит создать модель онтологии номинальных свойств, не зависящую от конкретной ПрО/ПрПО, т.е. позволит измерить близость (подобие) свойств объектов разных ПрПО.

Для решения поставленной задачи к структурным метаданным номинальных свойств отнесем, например, количества пробелов в значениях свойств, прописных букв, знака пунктуации «.», знака пунктуации «,», знака пунктуации «-»; наличие аббревиатур; наличие кавычек; часть речи и т.д.

На следующем шаге необходимо обработать каждое номинальное свойство каждого объекта потенциально подобных ПрПО, заполнив индивиды (экземпляры) концептов (классов) онтологии. Мощность множества концептов онтологии будет равна количеству номинальных свойств, участвующих в сопоставлении. Мощность множества индивидов в каждом концепте онтологии будет равна мощности множества кортежей соответствующего свойства.

Далее предлагается сопоставлять попарно концепты онтологии двух сравниваемых ПрПО d_i и d_j в порядке, определенном на этапе ранжирования объектов и их свойств методами, представленными ниже. При выявлении низкой степени соответствия, сравнить текущий концепт ПрПО d_i , со следующим по рангу концептом ПрПО d_j для исключения ошибки в выборе ранга свойства. Если же и на этом этапе мера подобия низкая, то исключить данные свойства из рассмотрения.

После отбора свойств с высокой и средней мерой подобия, анализировать их с привлечением экспертов, поскольку программный метод не дает гарантии исключения ошибок при анализе номинальных свойств, но значительно сократит время работы экспертов и снизит вероятность ошибок 1-го рода.

Методы обработки концептов онтологии

Для определения меры подобия свойств необходимо обработать индивиды концептов онтологии и составить таблицу сопряженности (или кросстаблицу) по следующим признакам:

Количество пробелов = 0; $0 < \text{Количество пробелов} < 3$; Количество пробелов ≥ 3 ; Количество прописных букв = 0; $0 < \text{Количество прописных букв} < 4$; Количество прописных букв ≥ 4 ; Количество знака пунктуации «.» = 0; $0 < \text{Количество знака пунктуации «.»} < 3$; Количество знака пунктуации «.» ≥ 3 ; Количество знака пунктуации «,» = 0; Количество знака пунктуации «,» = 1; Количество знака пунктуации «,» > 1 ; Количество знака пунктуации «-» = 0; Количество знака пунктуации «-» ≥ 1 ; Наличие аббревиатур = «есть»; Наличие аббревиатур = «нет»; Наличие кавычек = «есть»; Наличие кавычек = «нет»; Часть речи = «имя существительное»; Часть речи = «имя прилагательное»; Часть речи = «глагол». Таблица сопряженности или кросстабуляции – это таблица совместного распределения частот двух и более номинативных признаков, измеренных на одной группе объектов [9].

Мера сходства между концептами онтологии, а, соответственно, и между номинальными свойствами по комплексу признаков, анализировалась следующими методами: таксономический анализ Е. С. Смирнова [10], коэффициент взаимной сопряженности Пирсона [11], критерий согласия χ^2 Пирсона [12].

Таксономический анализ Е. С. Смирнова. В таксономическом анализе Е.С. Смирнова предполагается, что вес модальностей признаков различен в зависимости от частот их встречаемости. Чем реже встречается модальность в выборке, тем её вес больше и наоборот. При этом различают веса по присутствию и отсутствию одной и той же модальности. Следовательно, учитываются совпадения не только по присутствию тех или иных модальностей признаков, но и по их отсутствию. Всякому

несовпадению двух объектов по модальностям приписывается один и тот же вес «-1». Коэффициент сходства T_{ij} между i -ми концептами онтологий ПрПО d_i и d_j равен

$$T_{ij} = \frac{1}{M} \sum_{k=1}^M w_k, \text{ где } M - \text{общее количество модальностей по всем признакам; } w_k - \text{вес } k -$$

й модальности либо по присутствию ее, либо по отсутствию, либо по несовпадению их.

Вес по присутствию k -ой модальности w_k^+ , определяют по формуле: $w_k^+ = n_k / N_k$, а вес по отсутствию w_k^- : $w_k^- = N_k / n_k$, где N_k – число концептов, у которых данная модальность присутствует; n_k – число концептов, у которых модальность отсутствует.

Коэффициенты взаимной сопряженности Пирсона и Чупрова. Рассмотрим метод расчета коэффициента взаимной сопряженности Пирсона K_{Π} . Коэффициент взаимной сопряженности Пирсона применяется, если необходимо оценить тесноту связи между неальтернативными признаками (т.е. имеющими более 2-х групп в таблице сопряженности), которые могут принимать любое число вариантов значений; $K_{\Pi} = \sqrt{\varphi^2 / (1 + \varphi^2)}$, где φ^2 — показатель взаимной сопряженности, φ определяется как сумма отношений квадратов частот каждой клетки таблицы к произведению итоговых частот соответствующего столбца и строки. Вычитая из этой суммы 1, получим величину φ^2 . Чем ближе коэффициент взаимной сопряженности Пирсона к 1, тем связь показателей больше: $\varphi^2 = \sum n_{xy}^2 / (n_x n_y) - 1$.

Наряду с коэффициентом взаимной сопряженности Пирсона для определения тесноты связи номинальных признаков используют аналогичный коэффициент Чупрова, который рассчитывается по следующей формуле: $K_{\text{Ч}} = \sqrt{\varphi^2 (K_1 - 1)^{-1} (K_2 - 1)^{-1}}$, где K_1 и K_2 – число групп первого и второго признака соответственно. Интерпретация результата расчета коэффициента сопряженности Чупрова аналогична коэффициенту Пирсона, чем ближе значение коэффициента к 1, тем больше связь показателей.

Критерий согласия χ^2 -квадрат Пирсона. Критерий согласия хи-квадрат Пирсона – это непараметрический метод, который позволяет оценить значимость различий между фактическим (выявленным в результате исследования) количеством исходов или качественных характеристик выборки, попадающих в каждую категорию, и теоретическим количеством, которое можно ожидать в изучаемых группах при справедливости нулевой гипотезы. То есть, метод позволяет оценить статистическую значимость различий двух или нескольких относительных показателей (частот, долей).

Для расчета критерия χ^2 необходимо рассчитать ожидаемое количество наблюдений для каждой из ячеек таблицы сопряженности путем перемножения сумм рядов и столбцов с последующим делением полученного произведения на общее число наблюдений. Значение критерия χ^2 находится по следующей формуле:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c (O_{ij} - E_{ij})^2 / E_{ij}, \text{ где } i - \text{номер строки (от 1 до } r), j - \text{номер столбца (от 1 до } c),$$

O_{ij} – фактическое количество наблюдений в ячейке ij , E_{ij} – ожидаемое число наблюдений в ячейке ij . Для принятия решения о зависимости признаков необходимо сравнить значение расчетного критерия χ^2 с критическим значением при числе степеней свободы f . Число степеней свободы рассчитывается по формуле: $f = (r - 1) \times (c - 1)$. Критическое значение критерия χ^2 -квадрат Пирсона определяется по таблице. В том случае, если полученное значение критерия χ^2 больше критического, гипотеза H_0 отклоняется и принимается альтернативная H_1 . В качестве

гипотезы H_0 как правило принимается наиболее вероятный исход. В поставленной задаче предполагается, что свойства подобны, соответственно гипотезы имеют следующие значения: H_0 – свойства подобны; H_1 – свойства различны.

Апробация предложенных методов

Проверим работоспособность предложенных методов на примерах.

Предположим, дано 4 номинальных свойства: ФИО₁, ФИО₂, Город, Организация. Не будем учитывать ПрПО, объекты и ранги свойств для того, чтобы проверить качество работы предложенного подхода, сравнив заведомо подобные и разные свойства. Заполним индивиды (экземпляры) концептов (классов) онтологии по представленным свойствам. Построим таблицу сопряженности (кросстаблицу), подсчитав количество индивидов концепта онтологии, подходящих под определенный признак таблицы сопряженности.

Таксономический анализ Е. С. Смирнова. Определим меру сходства между концептами с помощью таксономического анализа Е. С. Смирнова. Выполним полный перебор концептов для определения максимальной и минимальной меры подобия. Признаки, по которым значения совпадают, отбрасываются. Отметим, что в данном методе расчет модальностей для двух концептов не целесообразен в силу своей специфики. Коэффициенты сходства T_{ij} представлены в табл. 1.

Таблица 1.

Коэффициенты сходства концептов

ФИО ₁ и ФИО ₂	0.0000
ФИО ₁ и Город	-0.3333
ФИО ₁ и Организация	-0.5333
ФИО ₂ и Город	0.0000
ФИО ₂ и Организация	-0.6000
Город и Организация	-0.5333

Проанализировав полученные результаты, можно сделать вывод, что наибольшее сходство между собой имеют свойства ФИО₁ и ФИО₂; ФИО₂ и Город. Последнее, как видно, ложно. Наименьшее сходство имеют свойства ФИО₂ и Организация; ФИО₁ и Организация; Город и Организация. Отметим, что мера подобия свойств ФИО₁ и Организация; ФИО₂ и Организация близки, что можно отнести к плюсам метода.

Коэффициенты взаимной сопряженности Пирсона и Чупрова. Рассчитаем коэффициенты взаимной сопряженности Пирсона K_{Π} и Чупрова $K_{\text{Ч}}$. Строки со значениями признаков равными нулю, отбрасываем. Для удобства расчетов разместим данные в таблице, просуммировав количество значений признаков по строкам и столбцам. Затем рассчитаем частное от деления квадрата значения каждого признака на произведение сумм в соответствующие строке и столбце.

Для нахождения φ^2 необходимо просуммировать полученные итоги и вычесть 1:

$$\varphi^2 = 0.5019 + 0.5019 - 1 = 0.0038, K_{\Pi} = \sqrt{0.0038 / (1 + 0.0038)} = 0.0614,$$

$$K_{\text{Ч}} = \sqrt{0.0038 / ((11 - 1)(2 - 1))} = 0.0195.$$

Рассчитаем коэффициенты взаимной сопряженности Пирсона и Чупрова для других пар концептов. Для анализа результатов сведем результаты расчетов в табл. 2.

Таблица 2.

Коэффициенты взаимной сопряженности Пирсона и Чупрова

Свойства	K_{Π}	$K_{\text{ч}}$
ФИО ₁ и ФИО ₂	0.0614	0.0195
ФИО ₁ и Город	0.5360	0.1914
ФИО ₁ и Организация	0.4635	0.1351
ФИО ₂ и Город	0.5319	0.2093
ФИО ₂ и Организация	0.4652	0.1357
Город и Организация	0.4912	0.1456

Коэффициенты у пар свойств ФИО₁ и Город, ФИО₂ и Город приблизительно равны, что подтверждает не случайность полученного результата. Эти же пары свойств имеют наиболее высокие коэффициенты. Наименьшие коэффициенты у пары свойств ФИО₁ и ФИО₂, причем значительно отличающиеся от коэффициентов других пар. Согласно интерпретации результатов, чем ближе коэффициенты к 1, тем выше зависимость признаков. Для поставленной задачи коэффициенты взаимной сопряженности Пирсона и Чупрова дают ложный результат.

Критерий согласия хи-квадрат Пирсона. Определим меру сходства между концептами с помощью критерия согласия χ^2 Пирсона.

Рассчитаем ожидаемые значения признаков, предварительно отбросив строки с нулевыми значениями признаков.

Определим значение χ^2 согласно формуле:

$$\chi^2 = \frac{50-50}{50} + \frac{50-50}{50} + \frac{49-49.5}{49.5} + \dots + \frac{50-50}{50} + \frac{50-50}{50} + \frac{50-50}{50} = 3.0303.$$

Число степеней свободы будет $(11-1)(2-1)=10$. Критическое значение χ^2 распределения при уровне значимости $\alpha = 0.05$ составляет 18.307. Сравнив полученное значение χ^2 распределения и критическое, можем сделать вывод о том, что гипотеза H_0 о подобии концептов ФИО₁ и ФИО₂ принимается с вероятностью статистической ошибки первого рода не более 5%.

Выполним расчеты для других пар концептов и результаты сведем в табл. 3.

Таблица 3.

Расчетные и критические значения χ^2 распределения

Концепты	$\chi^2_{\text{расч}}$	$\chi^2_{\text{крит}}$	Число степеней свободы
ФИО ₁ и ФИО ₂	3.0303	18.307	10
ФИО ₁ и Город	282.1724	19.675	11
ФИО ₁ и Организация	218.9259	24.996	15
ФИО ₂ и Город	276.1068	16.919	9
ФИО ₂ и Организация	220.9956	24.996	15
Город и Организация	222.5498	24.996	15

При сравнении других пар свойств полученное значение χ^2 распределения значительно превышает критическое значение, следовательно, гипотеза H_0

отклоняется, исходя из чего можем сделать вывод о том, что это абсолютно разные свойства, поскольку наблюдаемые и ожидаемые частоты значительно отличаются.

Метод «Анализ соответствий». Наиболее информативным из предложенных методов оказался критерий согласия хи-квадрат Пирсона. На основе этого критерия разработан метод «Анализ соответствий».

Анализ соответствий – это разведочный метод анализа, позволяющий визуально и численно исследовать структуру таблиц сопряженности большой размерности [13].

Строки и столбцы исходной таблицы представляются точками пространства, между которыми вычисляется расстояние хи-квадрат. Далее требуется найти пространство небольшой размерности, в котором вычисленные расстояния минимально искажаются, и в этом смысле максимально точно воспроизвести структуру исходной таблицы с сохранением связей между признаками.

Сравним концепты онтологии используя функцию программного обеспечения STATISTICA компании StatSoft [14].

Подготовим данные аналогично предыдущим методам. Исключим из таблицы сопряженности строки со значениями признаков равными «0» по обоим концептам. Создадим таблицу данных в программе или импортируем подготовленную заранее, например, из Excel. На вкладке «Анализ» выберем «Многомерный анализ -> Анализ соответствий». В появившемся окне в качестве входных данных выберем пункт «Частоты без группирующих переменных», т.к. исходные данные представлены таблицей сопряженности, и нажмем кнопку «ОК». На экране появится окно с кратким результатом.

Согласно методу анализа соответствий свойств ФИО₁ и ФИО₂ значение p -value значительно больше 0.05, следовательно, гипотезу H_0 о подобии свойств следует принять. Все другие пары имеют значение p -value равное 0, что говорит о том, что гипотезу H_0 следует отклонить, т.е. среди свойств подобие не найдено.

Вывод

Исходя из результатов тестирования предложенных методов для сопоставления номинальных свойств объектов различных предметных подобластей предлагается использовать модель онтологии и метод «Анализ соответствий» в комплексе с подходами, предложенными в технологии, описанной в [2]. Но для уточнения полученных результатов необходимо провести дополнительные исследования на других номинальных свойствах. Возможно, для повышения достоверности данных, необходимо для сопоставления свойств применять несколько предложенных методов.

Применение предложенного подхода автоматизирует поставленную задачу, что значительно сократит время работы для лица, принимающего решение, и снизит вероятность ошибок 1-го рода в отличие от автоматического решения задачи.

Список литературы

1. Малахов, Е.В. Оценка степени адекватности баз данных как информационных моделей предметных областей / Е.В. Малахов // Тр. Одес. политехн. ун-та.– 2004.– Вып. 1(21).– С. 82–86.
2. Glava, M. Searching Similar Entities in Models of Various Subject Domains Based on the Analysis of Their Tuples / M. Glava, E. Malakhov // 2016 International Conference on Electronics and Information Technology (EIT'16), May 23—27, 2016, Odesa, Ukraine, 2016. - pp. 97–100. ISBN 978-1-5090-2224-3 (DOI: 10.1109/ICEAIT.2016.7501001; EID: 2-s2.0-84979503116).
3. Колесникова, С.И. Методы анализа информативности разнотипных признаков / С.И. Колесникова // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. – 2009. – №1(6). – С. 69–80.

4. Никоненко, А.А. Обзор баз знаний онтологического типа / А.А. Никоненко // Штучний інтелект. — 2009. — № 4. — С. 208-219.
5. Бухановский, А.В. Метаонтология исследовательского проектирования морских динамических объектов / А.В. Бухановский, Ю.И. Нечаев // Онтология проектирования. — 2012. — № 1. — С. 53–63.
6. Полетаева, Е.В. Принципы построения онтологии предметной области машиностроения [Электронный ресурс] / Е.В. Полетаева // Электронный научный журнал: Программные продукты, системы и алгоритмы.— 2015.— № 1. — С. 1–3 Режим доступа: <http://swsysweb.ru/ontology-building-mechanical-engineering.html>.
7. Палагин, А.В. Онтологические методы и средства обработки предметных знаний: монография / А.В. Палагин, С.Л. Крытый, Н.Г. Петренко. — Луганск: ВНУ им. В. Даля. — 2012. — 324 с.
8. Тузовский, А.Ф. Системы управления знаниями (методы и технологии) / А.Ф. Тузовский, С.В. Чириков, В.З. Ямпольский.— Под общ. ред. В.З. Ямпольского. — Томск: Изд-во НТЛ, 2005.— 260 с.
9. Наследов, А.Д. Математические методы психологического исследования. Анализ и интерпретация данных / А.Д. Наследов.— Учебное пособие.— СПб.: Речь, 2007.— 36 с.
10. Марков, А.А. Методика сравнительного анализа информационных объектов на базе экспертных методов многокритериальной оценки: автореф. дис. ... к-та тех. наук : 05.13.01 / Марков, А. А. — М., 2003.— 166 с.
11. Громыко, Г.Л. Теория статистики: Учебник / Г.Л. Громыко.— Т11, 2-е изд., перераб. и доп.— М.: ИНФРА-М. — 2005.— 476 с.
12. Лапач, С.Н. Статистика в науке и бизнесе / С.Н. Лапач, А.В. Чубенко, П.Н. Бабич. — Киев: Морион, 2002.— 640 с.
13. Вуколов, Э.А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL: учебное пособие / Э.А. Вуколов. — 2-е изд.— М.: Форум, 2008.— 464 с.
14. Официальный сайт компании StatSoft [Электронный ресурс].— Режим доступа: <http://www.statsoft.com>.

ПОРІВНЯННЯ ВЛАСТИВОСТЕЙ НОМІНАЛЬНОГО ТИПУ ОБ'ЄКТІВ РІЗНИХ ПРЕДМЕТНИХ ПІДОБЛАСТЕЙ В РЕЛЯЦІЙНИХ БАЗАХ ДАНИХ

М.Г. Глава

Одесский национальный политехнический университет,
пр. Шевченко, 1, Одеса, 65044, Україна; e-mail: glavamaria@mail.ru

Розглядається проблема об'єднання моделей предметної області (ПрО). Пропонується зіставляти об'єкти ПрО на основі значень властивостей екземплярів цих об'єктів. Методи зіставлення властивостей розрізняються залежно від типу шкал, в яких вимірюються їх значення. Пропонується побудувати онтологічну модель для властивостей номінального типу і обробити її концепти за допомогою побудови таблиці спряженості ознак і методу аналізу відповідностей.

Ключові слова: модель предметної області, предметна підобласть, властивості номінального типу, таблиця спряженості ознак, аналіз відповідностей.

COMPARISON OF THE NOMINAL TYPE PROPERTIES OF OBJECTS OF DIFFERENT SUBJECT SUBDOMAINS IN RELATIONAL DATABASES

M. G. Glava

Odessa National Polytechnic University,
1, Shevchenko Ave., Odessa, 65044, Ukraine; e-mail: glavamaria@mail.ru

This article discusses the problem of the subject domain (SD) models merge. It is proposed to compare SD objects on the basis of the properties values of the tuples of these objects. The methods for comparing the properties of objects differ depending on the type of scales in which their values are measured. It is proposed to construct the ontological model for the nominal type properties and process its concepts by constructing contingency tables and correspondence analysis method.

Keywords: subject domain, model of subject domain, nominal type properties, contingency tables, correspondence analysis.