

МОНИТОРИНГ ДИНАМИКИ РАЗВИТИЯ ПРОЕКТА В ИНТЕЛЛЕКТУАЛЬНОМ ПРОЕКТНОМ РЕПОЗИТАРИИ

А.В. Чекина

ГОУ ВПО «Ульяновский государственный технический университет», г. Ульяновск

Рецензент В.Е. Подольский

Ключевые слова и фразы: версия проекта; индексирование; информационный ресурс; мониторинг динамики проекта; нейросетевая кластеризация; проектный репозиторий; частотный портрет; FCM кластеризация.

Аннотация: Описаны возможности мониторинга динамики развития проекта в интеллектуальном репозитории, применительно к информационным ресурсам, которые могут быть представлены для анализа своими частотными портретами. Представлены результаты экспериментов кластеризации таких информационных ресурсов.

Введение

Современный проектный репозиторий должен представлять собой интеллектуальное хранилище информационных ресурсов, чтобы обеспечить поиск необходимого ресурса на основе гибкого запроса. Особенность проектных репозитариев, отличающая его от универсального информационного хранилища, – это учет хранения многих версий одного и того же информационного ресурса. Основу индексирования информационных ресурсов традиционно составляет лексический портрет текстового дескриптора ресурса. Однако для организации интеллектуального проектного репозитория индексирования на основе частотных словарей терминов недостаточно, необходимо индексирование по времени создания и версиям.

По мере развития любого проекта появляются многочисленные версии одного и того же документа. Кластеризация проекта в разрезе версий позволит получить дополнительную информацию для проектного менеджмента. В каждый период времени проект характеризуется разной картой кластеризации составляющих его документов. На основе динамической кластеризации проекта может быть построена эффективная система мониторинга проектов.

Чекина А.В. – аспирант кафедры «Информационные системы» УлГТУ, г. Ульяновск.

1. Формализованная постановка задачи нечеткой кластеризации

Пусть исследуемая совокупность представляет собой конечное множество элементов $A = \{a_1, \dots, a_n\}$, которое получило название множество объектов кластеризации. В рассмотрение вводится конечное множество признаков или атрибутов $P = \{p_1, \dots, p_q\}$, каждый из которых количественно представляет некоторое свойство или характеристику элементов рассматриваемой проблемной области. При этом n – общее количество объектов данных, а q – общее количество измеримых признаков.

Далее предполагается, что для каждого из объектов кластеризации некоторым образом измерены все признаки множества P в некоторой количественной шкале. Тем самым, каждому из элементов $a_i \in A$ поставлен в соответствие некоторый вектор $x_i = (x_1^i, x_2^i, \dots, x_q^i)$, где x_j^i – количественное значение признака $p_j \in P$ для объекта $a_i \in A$. Для определенности будем предполагать, что все x_j^i принимают действительные значения.

Векторы значений признаков $x_i = (x_1^i, x_2^i, \dots, x_q^i)$ удобно представлять в виде матрицы данных D размерности $n \times q$, каждая строка которой равна значению вектора x_i .

Задача нечеткого кластерного анализа формулируется следующим образом: на основе исходных данных D определить такое нечеткое разбиение $R(A) = \{A_k \mid A_k \subseteq A\}$ или нечеткое покрытие $J(A) = \{A_k \mid A_k \subseteq A\}$ множества A на заданное число c нечетких кластеров $A_k, k \in \{2, \dots, c\}$, которое доставляет экстремум некоторой целевой функции $f(R(A))$ среди всех нечетких разбиений или экстремум целевой функции $f(J(A))$ среди всех нечетких покрытий.

Для решения задачи требуется дополнительно уточнить вид целевой функции и тип искоемых нечетких кластеров (поиск нечеткого разбиения или покрытия).

2. Виды нечеткой кластеризации

В настоящее время разработаны кластеризации следующих видов:

- частично-управляемая (partial supervision);
- на основе сходства (proximity clustering);
- на основе контекста (context based clustering);
- на основе сотрудничества (collaborative clustering).

Не все из перечисленных методов пригодны для построения системы управления проектным репозитарием. Необходимым свойством кластеризации должна быть возможность применения к вычислимым индикаторам встречаемости терминов.

3. Структурно-функциональное решение интеллектуального хранилища

Программная система, реализующая идеи интеллектуального хранилища, состоит из следующих подсистем: подсистема индексирования электронного информационного ресурса (ЭИР) (индексатор), подсистема кластеризации ЭИР на основе нейронной сети (нейросетевой кластеризатор) [1–3], подсистема кластеризации на основе Fuzzy C-Means метода (FCM-кластеризатор) и классификатор (рис. 1).

На модуль индексации возложены задачи предобработки текстовых документов или аннотаций к ЭИР и построение частотных словарей встречающихся терминов. Сохранение частотных таблиц необходимо производить в СУБД MS SQL 2000. Далее, в рамках модуля кластеризации и классификации, на основе значений относительных частот должны создаваться предметно-ориентированные кластеры, которые организуются в виде иерархии. В процессе классификации выполняется задача соотнесения вновь заносимого ЭИР с определенным кластером.

4. Результаты экспериментов

ФНПЦ ОАО «НПО Марс» (г. Ульяновск) является крупной проектной организацией с давней историей и обладает значительным архивом успешных проектов. Новые проекты должны использовать ранее разработанные решения, так как повторность использования позволяет сократить сроки проектирования.

Документы похожи, если кластеризатор отнес их в один кластер. С другой стороны, документы, являющиеся версиями из одного проекта, могут быть отнесены к разным кластерам, если в какой-то момент были приняты кардинально новые проектные решения. Поэтому мониторинг версий

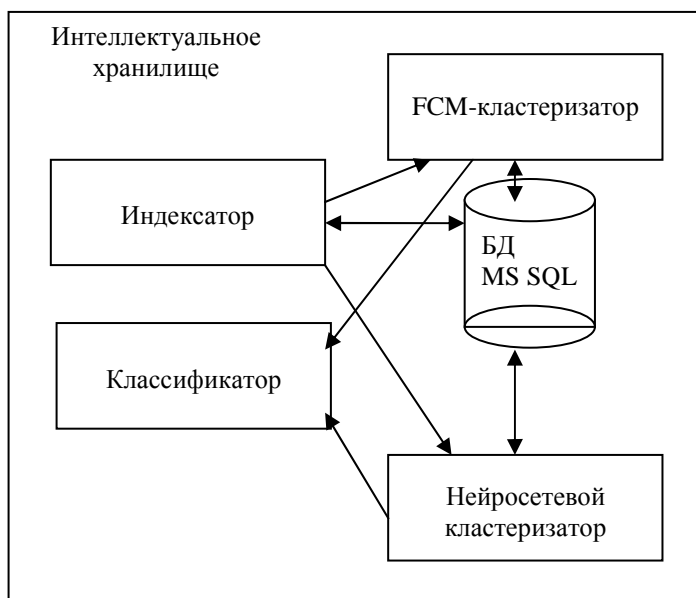


Рис. 1. Структура интеллектуального хранилища

Таблица 1

| |
|--|
| ИНФОРМАЦИОННАЯ СИСТЕМА ПРЕДПРИЯТИЯ. WEB-СЕРВЕР ПРЕДПРИЯТИЯ. Руководство пользователя. Инструкция пользователю АБЕИ.97071.007 ИЭ-1.doc |
| ИНФОРМАЦИОННАЯ СИСТЕМА ПРЕДПРИЯТИЯ. WEB-СЕРВЕР ПРЕДПРИЯТИЯ. Руководство пользователя. Инструкция администратору АБЕИ.97071.007 ИЭ-2.doc |
| ИНФОРМАЦИОННАЯ СИСТЕМА ПРЕДПРИЯТИЯ. WEB-СЕРВЕР ПРЕДПРИЯТИЯ. Программа и методика испытаний АБЕИ.97071.007 ПМ.doc |
| Перечень средств вычислительной техники |
| Автоматизированное рабочее место. Система локальной электронной почты. Руководство пользователя. Часть 1. Настройка серверной части АБЕИ.97071.009 ИЗ-1.doc |
| Автоматизированное рабочее место. Система локальной электронной почты. Руководство пользователя. Часть 2. Настройка клиентской части АБЕИ.97071.009 ИЗ-2.doc |

Таблица 2

| Документ | Кластер №1 | Кластер №2 | Кластер №3 | Кластер №4 | Кластер №5 |
|-------------------------|------------|------------|------------|------------|------------|
| АБЕИ.97071.007 ИЭ-1.doc | 16,70 | 17,15 | 15,94 | 16,08 | 18,96 |
| АБЕИ.97071.007 ИЭ-2.doc | 16,66 | 17,06 | 16,17 | 16,26 | 18,13 |
| АБЕИ.97071.007 ПМ.doc | 16,76 | 17,03 | 16,02 | 16,21 | 18,69 |
| АБЕИ.97071.007 С2.doc | 16,72 | 17,07 | 16,01 | 16,14 | 18,79 |

документов, построенный на основе кластеризации, может служить эффективным инструментом проектного мониторинга. Рассмотрим возможности такого мониторинга на конкретном примере.

Фрагмент исходного множества документов представлен в табл. 1. Результат кластеризации приведен в табл. 2.

Заключение

Разработанная система на основе нечеткой кластеризации может эффективно применяться в интеллектуальных проектных репозиториях для мониторинга динамики развития проекта.

Список литературы

1. Ярушкина, Н.Г. Основы теории нечетких и гибридных систем / Н.Г. Ярушкина. – М. : Финансы и статистика, 2004. – 320 с.
2. Нечеткие гибридные системы. Теория и практика / И.З. Батыршин [и др.] / под ред. Н.Г. Ярушкиной. – М. : Физматлит, 2007. – 208 с.
3. Наместников, А.М. Интеллектуальный сетевой архив электронных информационных ресурсов / А.М. Наместников, А.В. Чекина, Н.В. Корюнова // Программные продукты и системы. – 2007. – № 4. – С. 10–13.

4. G. Stumme A. Hotho, B. Berendt. Semantic Web Mining. State of the art and future directions/ Web Semantics: Science, Services and Agents on the World Wide Web. – 2006. – №4. – P. 124–143.

5. L. Stojanovic at al. The role of ontologies in autonomic computing systems // IBM Systems Journal. – 2004. – Vol. 43, №3. – P. 598–616.

6. Инмон, Б. DW 2.0: хранилища данных следующего поколения / Б. Инмон // Открытые системы. – 2007. – №5.

Monitoring of Project Advancement Dynamics in Intelligent Design Repository

A.V. Chekina

Ulyanovsk State Technical University, Ulyanovsk

Key words and phrases: project version; indexing; information resource; monitoring of project development; clustering based on neural networks; design repository; project repository; frequency portrait; FCM-clustering.

Abstract: The paper describes the possibilities of monitoring of project advancement dynamics in intelligent repository with regard for information resources which can be analyzed through their frequency portraits. The results of information resources clustering are presented.

© А.В. Чекина, 2008