

ния новой информации/корректировки, новые/измененные данные выделяются красным цветом.

Гл. табельщик. ИС автоматически сообщает о не подтвержденных/непроверенных графиках работы. Табельщик при помощи соответствующих инструментов утверждает расписание сотрудников. После утверждения расписание подсвечивается синим цветом, что в свою очередь является сигналом для руководителей подразделений и сигналом для службы безопасности (СБ), для внесения соответствующей информации в турникет.

Сотрудник СБ. При входе в ИС получает сведения о не сверенной с турникетом информации о графиках работы. Интерфейс аналогичен интерфейсу гл. табельщика.

Диспетчер. На основании информации о графиках работы пользователей, доступной ему для любого подразделения организует работу автотранспорта.

Дежурный СБ. На основании доступной в ИС информации осуществляет контроль по допуску сотрудников на предприятия, если это не согласуется с обычным графиком работы, например, в выходные и праздничные дни.

Сотрудники столовой. На основании доступной информации из ИС формируют потребность в приготовлении необходимого количества питания для сотрудников.

Данная ИС прошла успешную апробацию на протяжении более чем одного года, является унифицированной и может быть использована на др. подобных предприятиях.

Литература

1. Дейт К. Введение в системы баз данных, 6-е издание : Пер. с англ. – К.; М.; СПб.: Вильямс, 2000.-848с.

Кластеризация электронных ресурсов в хранилищах, библиотеках, фондах

О.Н. Половикова

АлтГУ, г. Барнаул

Определена проблема выражения и интерпретации поисковых потребностей пользователей электронных ресурсов в хранилищах, библиотеках, фондах. Для решения данной проблемы предложено изменить классическую схему поискового аппарата «запрос-ответ», на «запрос – тематический раздел – ответ». Построен механизм формирования тематических разделов на основе кластерного анализа для оптимизации поискового аппарата в справочно-информационных фондах.

Проведена сравнительная характеристика методов и алгоритмов кластеризации с перспективой их использования для создания тематических разделов.

Непрерывный рост базовых и дополнительных ресурсов в различных информационных хранилищах (фондах, библиотеках, архивах и т.д.) создал проблему предоставления пользователям этих систем эффективного поискового аппарата. В технологиях и методах создания информационных хранилищ произошло перераспределение акцентов, определяющих основные принципы их проектирования и реализации. Если раньше модель хранения информационных ресурсов определяла систему и аппарат поиска, то в современных информационных хранилищах модель поисковой системы (учитывающая потребности пользователей и свойства ресурсов) определяет принципы хранения и обработки ресурсов.

Проблема организации поиска является актуальной и для широкого используемых в различных предметных областях справочно-информационных фондов, поэтому качественно изменились требования к разработке для них эффективного поискового аппарата.

Справочно-информационный фонд (СИФ) – собрание источников информации определённой тематики, имеющее справочный аппарат для их поиска. СИФ комплектуется на основе изучения информационных потребностей той категории потребителей, которых они призваны обслуживать. В функции СИФ входят: сбор, обработка и хранение как опубликованных материалов, так и неопубликованной научно-технической документации (отчёты, проекты, рационализаторские предложения, депонированные рукописи и т. п.); поиск информации по определённому тематическому профилю; обеспечение предприятий, организаций и отдельных специалистов первичными и вторичными источниками информации, а также фактическими сведениями (фактографическая информация).

Определяя принципы справочного аппарата поиска для СИФ в первую очередь необходимо учитывать потребности будущих пользователей данной системы и ставшие классическими проблемы, связанные с интерпретацией условий поиска, формируемых пользователем. Выражая свои требования на поиск ресурсов, в условия недостаточного владения информацией в соответствующей области (сфере), а также без практического опыта работы с конкретной поисковой системой, пользователь не способен выразить, а главное передать системе свои информационные потребности. Формируя требования поиска, пользователь рискует пропустить важные для него источники, потенциально имеющие отношения к его обучению.

Информационный запрос может существенно отличаться от информационных потребностей, так как несколькими значениями информационных полей трудно выразить смысл, содержание и объем

интересующего материала. Кроме того, пользователь не всегда осознает, какая именно информация ему действительно нужна. Поэтому на этапе формулировки информационного запроса потребительские потребности существенно огрубляются.

Предлагаемый подход, основанный на использовании тематических разделов, позволяет направлять (координировать) пользователя в процессе поиска и избежать перечисленных недостатков классической технологии «запрос-ответ». Предоставляя пользователю, множество информационных ресурсов, относящихся к определенному тематическому разделу, система направляет поиск пользователя в информационную область, которая отвечает его потребностям.

Изменение классической схемы информационного обслуживания (общения пользователя с системой) «запрос – ответ» на схему «запрос – тематический раздел – поисковый образ ресурса – информационный ресурс» требует привлечения алгоритмов для построения тематических разделов. Разбиение ресурсов на тематические разделы следует организовать таким образом, чтобы близкие ресурсы по смыслу ресурсы принадлежали одному тематическому разделу (кластеру).

Постановки задач кластеризации, автоматической классификации, самообучения, таксономии базируются на процессах группировки, классификации объектов по схожести их свойств. Аппараты решения перечисленных задач имеют общие подходы, используют похожие алгоритмы. Таким образом, можно не конкретизировать решаемую задачу до определения её конкретного названия и использовать смешанный аппарат для её решения.

Одно и то же множество из объектов можно разбить на кластеры (таксоны, группы) разными способами. Выбор критерия группировки позволяет отличать «хорошую» группировку от «плохой» и останавливаться на «наилучшем» варианте кластеризации (таксономии). Критерий разделения должен соответствовать содержательному смыслу задачи классификации.

Постановка задачи

Задача классификации множества информационных ресурсов: необходимо определить технологию (метод) для оптимального разбиения (классификации) поисковых образов ресурсов на информационные разделы. Требуется классифицировать информационные ресурсы, т.е. представить множество ресурсов некоторым числом групп (таксонов). Количество таксонов должно принадлежать некоторому заданному отрезку (указаны возможные пределы).

Поисковый образ ресурса – это совокупность свойств ресурса, необходимых пользователю системы в процессе физического или логического поиска. Поисковый образ ресурса отражает только те признаки и характеристики реального ресурса, которые актуальны для данного справочно-информационного фонда.

При этом необходимо показать, что ресурсы, соответствующие одному таксону, были бы достаточно близки друг к другу с точки зрения некоторого критерия близости, а ресурсы из разных таксонов были бы достаточно далеки друг от друга по этому критерию.

Решение задачи

Решением задачи кластерного анализа является такое разбиение ресурсов на кластеры p_1, p_2, \dots, p_m , которое удовлетворяет некоторому математическому критерию качества кластеризации (критерию оптимальности). Этот критерий может быть записан в виде функционала, выражающего уровни желательности различных разбиений множества ресурсов на тематические разделы.

При разработке критерия качества кластеризации следует учитывать следующие неформальные требования:

1) внутри одного кластера объекты должны быть связаны между собой;

2) объекты из разных кластеров должны быть далеки друг от друга;

3) при прочих равных условиях распределение объектов по кластерам должно быть равномерно.

Важным свойством для кластера являются связность (близость) – эти свойства предоставляют основу для построения функций критерия качества разбиения. Для количественного измерения сходства двух объектов необходимо построить функцию меры близости двух объектов. От данной функции зависит выбор варианта разбиения объектов на кластеры при заданном алгоритме разбиения. Выбор варианта разбиения зависит от целей исследования, физической и статистической природы, используемой информации.

Согласно алгебраической структуре представления информационных разделов «близость» можно интерпретировать как совпадение значений информационных слоев, т.е. таких свойств ресурса, как ключевых слов. Чем больше совпадений, тем ресурсы «ближе» и наоборот. Степень близости двух ресурсов призвана определять функция близости, которая отражает число совпадений.

Для объектов, чьи признаки невозможно представить числовым вектором характеристик, определим такое свойство, как среднее сходство объектов кластера. Среднее сходство объектов кластера показывает, насколько сходны между собой объекты (насколько они похожи друг на друга, насколько объекты связаны).

Функция, отражающая среднее сходства объектов кластера, может быть построена как среднеарифметическая (или как среднеквадратичная) мера значений функции близости. Выполнение одного из неформальных требований кластеризации: внутри одного кластера объекты должны быть связаны между собой, можно проверить по значениям функции среднего сходства объектов для всех построенных кластеров.

На основании анализа методологических особенностей процессов кластеризации сформируем следующие шесть этапов решения задачи разбиения ресурсов на тематические разделы:

1. Отбор объектов для кластеризации (определение ресурсов, которые подлежат кластеризации).
2. Определение множества признаков, по которым будут оцениваться ресурсы в выборке – признаки близости.
3. Определение меры близости между объектами и вычисление её значений для выбранных ресурсов.
4. Построение критерия качества кластеризации.
5. Выбор и применение одного из методов (или комбинации методов) для создания тематических разделов сходных ресурсов по критерию качества кластеризации.
6. Проверка достоверности результатов кластерного решения.

Литература

1. Игнатов И.Г. Резонтов К.В. Создание и использование в репозитории карточек информационных ресурсов на основе стандарта метаданных // Труды X конференции Телематика'2003. – СПб., 2003. С. 175–176.
2. Загоруйко Н.Г. Прикладные методы анализа данных и знаний. – Новосибирск: Изд-во Института математики, 1999.
3. Ким Дж.-О., Мьюллер Ч.У., Клекка У.Р. и др. Факторный, дискриминантный и кластерный анализ.: Пер. с англ. – М.: Финансы и статистика, 1989.
4. Кузнецов С.Д. Введение в информационные системы // Системы управления базами данных. Издательство «Открытые системы», 1997. – № 2.

Разработка web-ресурса для информационной поддержки специалистов сельскохозяйственного производства

Е.В. Понькина, О.В. Русакова, С.А. Александров
Алтайский государственный университет

Эффективность сельскохозяйственного производства зависит от множества факторов, включающих климатические и почвенные условия, особенности используемой агротехнологии, наличие современной и адекватной используемым технологиям обработки почвы и природным особенностям производства сельскохозяйственной техники и оборудования, рациональной организации сельскохозяйственного произ-