

Table of Contents

Классификаторы - борьба с энтропией 2

Классификаторы - борьба с энтропией

- Справочник учебных заведений в Живом Журнале
- Справочник материально-технических ресурсов
- Классификатор предприятий и объектов энергетики
- Общероссийский классификатор видов деятельности, продукции и услуг
- Библиотека Мошкова

Структура кодов. Контрольные числа. Составные классификаторы (матричная классификация). Коды с внутренней иерархией (иерархическая классификация).

Проблемы - невозможность централизованного ведения классификатора (восходящая информация на местах) - множество точек ввода информации. Как следствие - дублирование информации. Высокая важность удобства инструментов поиска. Синонимы и закладки, дополнительные поля классификатора, облегчающие поиск. Борьба - экспертная команда "вычистки" справочников. Конфликт интересов - разработчиков инструмента и экспертной команды (на примере ЖЖ). Многоуровневая организация экспертов (модераторы и активисты), привлечение новых экспертов.

Необходимость углубления детализации справочника по разным ветвям - "перекос" дерева классификатора. Необходимость мотивации источников данных к построению общего классификатора.

Историческое изменение справочников. Примеры запросов к классификаторам. Множественные иерархии. Проблема отслеживания преемственности в условиях изменяемых атрибутов (преемственность организации, преемственность технических объектов, зданий, школ).

Распространение информации из центрального хранилища во вне - проблемы совместимости интерфейсов. Стандарты на формат обмена данными - обязательные (были во времена СССР) и необязательные (сегодня).

Проблема авторских прав и распространяемости классификатора. Риск, что классификатор станет частной закрытой базой данных. Модели заработка на поддержке и развитии классификатора. Свободные классификаторы.

Классическая модель хранилища данных (справочники в виде длинной "простыни") и объектный подход (машиночитаемый архив с "выписками" для человека). Исторические причины.

Альтернатива классификации - система тэгов (максимально матричная классификация). Классифицирование фотографий (поднять идею кто-где-когда).

Отличие классифицирования от идентифицирования на примере базы МТР по нескольким предприятиям, в которой с одной стороны решена проблема уникальной идентификации продуктов методом выдачи кодов предприятий, а потом уже предприятие может само управлять своими кодами продукции - известной международной системой штрихкодов продукции. И не решённая этой системой задача классификации, так как тяжело или невозможно не зная заранее найти аналоги продукции, замены и просто нужную позицию у произвольного предприятия.

Любой классификатор является по сути моделью мира, поэтому идеальный классификатор содержит в себе полную модель мира, однако это конфликтует с ограниченностью ресурсов на создание справочника и невозможность синхронизации в реальном времени с реальным положением дел. Чисто технически не получается сделать эту штуковину. Причём такие классификаторы практически становятся базами знаний по предметной области и отвечают на вопросы: “что является чем”, “что куда вставить”, “какой объект является частью другого объекта”, “что чем можно заменить”. Получается экспертная база, а не просто классификатор каких-то объектов.

Примеры видов иерархии в одном и том же классификаторе, например предприятий:

- Какое предприятие кем владеет - реестр предприятий с описанием долей
- Какое предприятие с кем имеет контракты и деловые связи (сеть)

При классифицировании обычно выделяют корень, то есть строят древовидную систему с единственным родительским элементом, от которого идут все остальные. Однако в жизни не всегда удаётся так сделать, потому что часто данные объекты взаимосвязаны по типу сети, невозможно выделить естественным образом единственный корень, есть кроссылки, петли и т.д., приводящие к усложнению структуры. Например классификатор сообществ Живого Журнала и пользователей в них.

From:

<https://kibi.ru/> - **КибИ.ру**

Permanent link:

<https://kibi.ru/notes/klassifikatory>

Last update: **2016/10/17 20:12**

