


«Автоматический авторитетный контроль»

Князева А.А., Колобов О.С.
ИВТ СО РАН, ИСЭ СО РАН
ООО «Комлог»



Сокращения

- * АК – авторитетный контроль
- * БЗ – библиографическая запись
- * АЗ – авторитетная/нормативная запись

Задача

При выполнении работ по развитию портала НП «МедАрт», появилась задача поиска работ конкретного автора в различных источниках библиографических данных.

Итак, у нас имеется -

*Один или несколько источников библиографических записей

*Один авторитетный файл «имя лица»

Точка доступа «автор»

- * Не всегда в полях данных есть расшифровка имени автора до полной формы
- * Не работает для случая, если автор меняет фамилию
- * Не работает для случая, когда имя автора может записываться с вариантами

Основные функции АК

- * АК. Обеспечение единообразия заголовков
- * Поиск. Обеспечение связи с различными формами заголовков
- * Информационная. Достоверные примечания и комментарии каталогизатора.
- * **Техническое обслуживание.
Выявление ошибок и корректировка библиографических записей**

Идея

- * Рассматривается соответствие пары записей АЗ-БЗ
- * Машинное обучение основанное на **обучающей выборке**
- * Введение классов соответствующих и несоответствующих пар записей АЗ-БЗ
- * Задание правил вычисления расстояния между парами записей АЗ-БЗ для определения близости к центроидам того или иного класса

Роль обучающей выборки

На основе обучающей выборки можно оценить значимость признаков. Чем важнее признак, тем больший вес он получит в процессе обучения.

Анализ базы данных «MedArt»

- * Библиографические записи аналитического уровня базы данных «MedArt»
- * Авторитетные записи базы данных «Медики России»
- * Для анализа рассматривалось 41746 упоминаний авторов

Основная информация

- * Профессия
- * Годы жизни автора
- * Имя и отчество в полной форме
- * Место работы автора
- * Номер авторитетной записи

Дополнительная информация

- * Группа 1. Информация о коллективном авторе
- * Группа 2. Информация о соавторах
- * Группа 3. Информация о предметных рубриках

Результат

При связывании АЗ-БЗ предпочтительно полагаться на –

- *Информацию о соавторах без кода авторитетной записи

- *Информацию о предметных рубриках

Технология. Основные этапы

- * Нормализация полей данных
- * Составление пар записей
- * Сравнение отдельных полей или подполей для пары записей
- * Принятие решения
- * Обучение решающей функции
- * Оценка качества связывания

Большее информации

- * Принципы идентификации объектов в структурированных документах / А.А. Князева // Вестник НГУ. Сер.: Информационные технологии. – 2013. – Vol.36, № 1. – С. 12-19.

«Автоматический авторитетный контроль»

Колобов О.С.
okolobov@gmail.com

Спасибо за внимание !