



ИНВЕСТИЦИОННОЕ ПРЕДЛОЖЕНИЕ

Система интеллектуальной классификации текстов

Возможные партнеры: различные консалтинговые компании, деятельность которых требует постоянного анализа больших объемов текстов на естественном языке и быстрого отбора релевантной (некоторым темам) текстов.

Необходимые инвестиции и сроки: 25 млн. рублей на 2 года.

Классификация текстов на естественном языке

Задача классификации текстов заключается в разделении документов на фиксированное число предопределенных категорий (классов). Такая проблема довольно часто встает, когда необходимо отыскать в большом документальном архиве документы, обладающие определенными содержательными характеристиками, рубрицировать или, что тоже часто требуется, перерубрицировать документы.

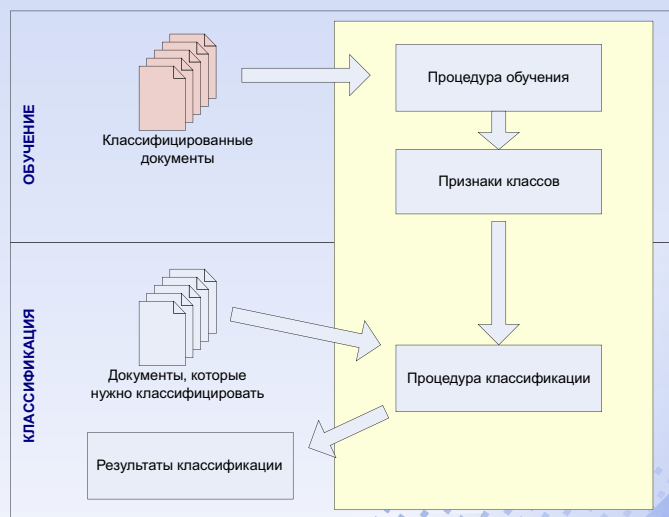
В ИПС разработана программная система АКТИС, позволяющая эффективно решать проблему классификации документов. Она легко настраивается на новые задачи классификации достаточно указать системе на несколько документов, классы которых известны, и она сама обучится распознавать в дальнейшем документы этих классов. Это особенно важно, если требуется часто перерубрицировать документы, или в случае, когда количество классов очень велико.

Данная система, в случае ее доработки до завершеного коммерческого продукта, могла бы найти применение в системах документооборота и делопроизводства в качестве средства для отбора (фильтрации) документов определенного рода, рубрикации документов, систематизации хранения документов, регулирования потоков документов.

Решения для классификации: особенности и преимущества

- ✦ Признаки классов — не просто отдельные слова (лексика), но и словосочетания (и другие более сложные синтаксические конструкции), которые позволяют точнее отразить смысл текста (а значит и точнее определить класс документа).
- ✦ Архитектура программного средства позволяет использовать преимущества кластерной аппаратной платформы (с очень высокой степенью утилизации процессорных ресурсов), что делает возможным создание высокопроизводительных масштабируемых ресурсов классификации текстовой информации.

Общая схема работы системы





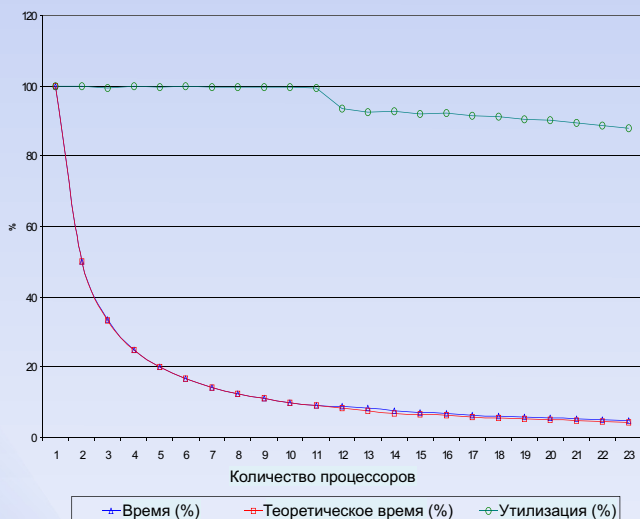
Возможный сценарий коммерциализации

Для данной программной системы видится целесообразным следующий сценарий коммерциализации.

- ✦ Доработка программного продукта по следующим направлениям:
 - ✦ разработка удобного визуального пользовательского интерфейса;
 - ✦ разработка конфигурационных средств и организационных сценариев для использования системы классификации в варианте ее интеграции с другими программными средствами (в том числе, интеграция в системы документооборота);
 - ✦ оптимизация алгоритмов до промышленного уровня.
- ✦ Патентование методологических решений.
- ✦ Создание схемы продвижения продукта и группы сопровождения.
- ✦ Реализация.

Для данного сценария приемлемо как инвестирование вновь созданного предприятия, так и создание совместного предприятия, в котором долей ИПС будут являться научно-технические разработки и программные решения. Инженерно-технический кадровый состав ИПС способен осуществить предложенную доработку программного продукта в течение полутора-двух лет.

Степень утилизации вычислительных ресурсов в зависимости от количества вычислительных узлов



Статусное окно процесса классификации

