HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

PROMT
TRANSLATION SOFTWARE

Использование машинного перевода в системах поиска русскоязычной информации

RUSSIAinfo, университет Хельсинки

Компания ПРОМТ, Санкт Петербург

Таня Пурсиайнен

Дарьяна Цугульская

- Информационная служба RUSSIAinfo создана в университете Хельсинки по инициативе и при поддержке Министерства Просвещения Финляндии
- RUSSIAinfo предлагает доступ к электронным ресурсам по России для академического международного круга пользователей
- Реферативная база данных
- Метаданные: на английским и на финском
- Поиск: на английским и на финском
- Ресурсы: 32% на английском языке, 12% на финском языке, 55% на русском языке, 11% на других языках

- Для пользователей, не владеющих русским языком, RUSSIAinfo предлагает возможность машинного перевода текстов с русского на английский (лицензия от ПРОМТ)
- В настоящий момент предлагается три способа использования машинного перевода: перевод текста, перевод URL, и автоматический перевод результата поиска
- Цель подключения МП – повышения коэффициента полноты поисковой системы: русскоязычные ссылки не отбрасываются пользователем
- Качество перевода: самое высокое достигается при переводе текстов по экономической тематике (специальные словари)
- Сотрудничество с компанией ПРОМТ позволит нам также повысить качество перевода текстов по другим тематикам

- Машинный перевод: черновой вариант перевода
- Быстрый перевод текста с целью понять смысл
- Полное соответствие идеологии WWW: пользователь привык быстро «просматривать» веб-страницы и немедленно получать информацию
- Лингвистическая база, программная база. Общелексические и специальные словари.
- Позволит решить одну из основных проблем многоязычного поиска информации: перевод поисковых выражений (фраз)

- От чего зависит точность перевода?

  - Грамотность исходного текста: Грамматика, правописание

  - Наличие слов в переносном значении, неологизмов, аббревиатур

- Улучшение качества перевода специализированных текстов

- Подключение специализированных словарей (Созданных ПРОМТ / Созданных пользователем)

- Создание списка зарезервированных слов

- Пример: Перевод документации по теме стоматология

  1. С использованием созданного компанией ПРОМТ специализированного словаря
  2. С использованием общелексического словаря

Example:

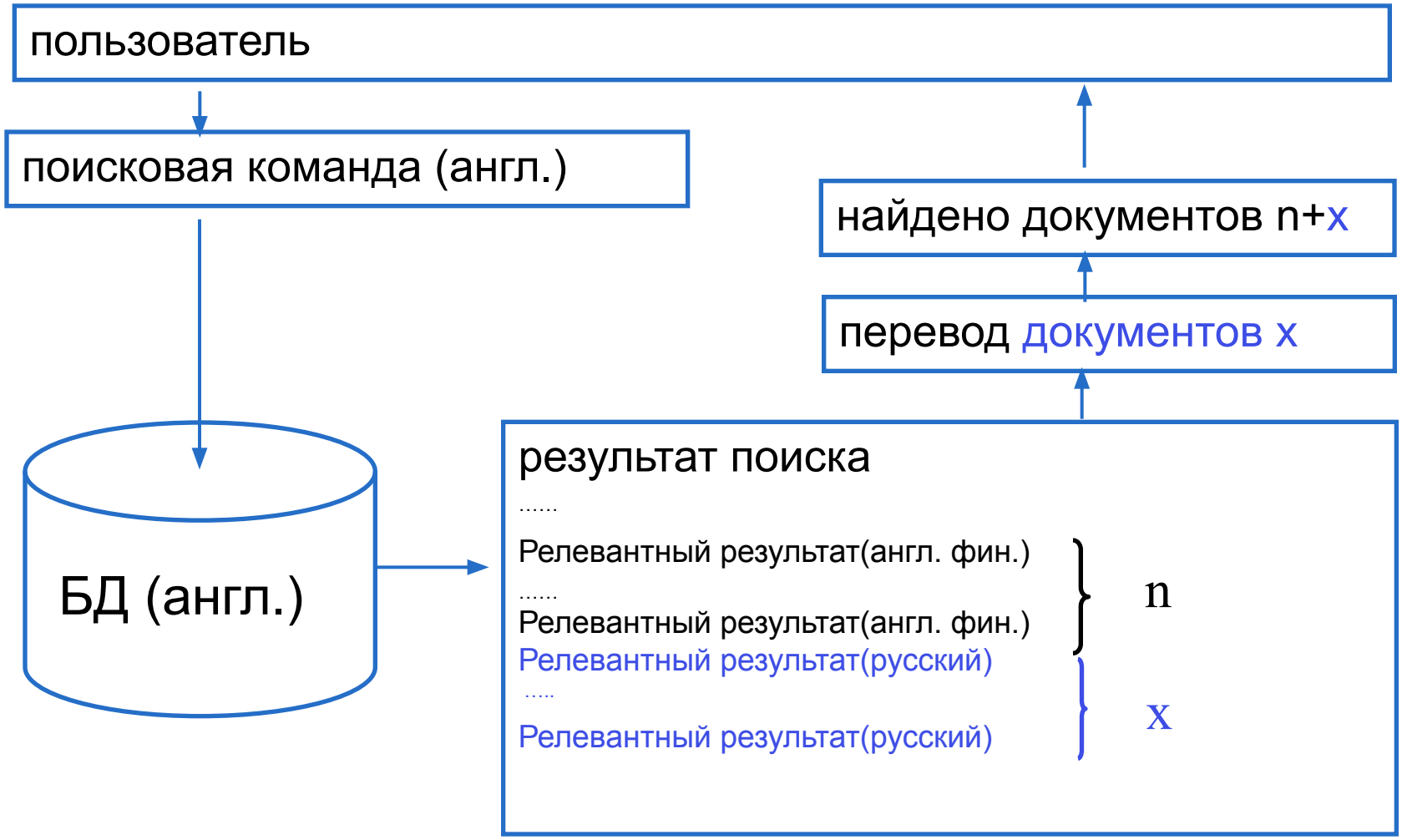
The layer of material beneath tooth enamel is the dentine. It too is composed of hydroxyapatite to the extent of about 70 per cent, the remainder is collagen and water. The dentine matrix is perforated by a number of tiny canals which radiate from the pulp cavity to the surface. These are the dentine tubules.

Перевод с использованием специализированного словаря, созданного на заказ:	**Перевод с использованием общелексического словаря системы:**
Слой материала под эмалью зуба - дентин. Это также составлено из гидроксиапатита вплоть до, приблизительно 70 процентов, остаток - коллаген и вода. Матрица дентинов перфорирована множеством крошечных каналов, которые исходят от полости зуба до поверхности. Они - зубные канальцы.	Слой материала ниже эмали зуба - dentine. Это также составлено из hydroxyapatite вплоть до приблизительно 70 процентов, остаток - collagen и вода. dentine матрица перфорирована множеством крошечных каналов, которые исходят от впадины целлюлозы до поверхности. Они - dentine tubules.

- Создание специализированных словарей для RUSSIAinfo
- Российские государственные учреждения:

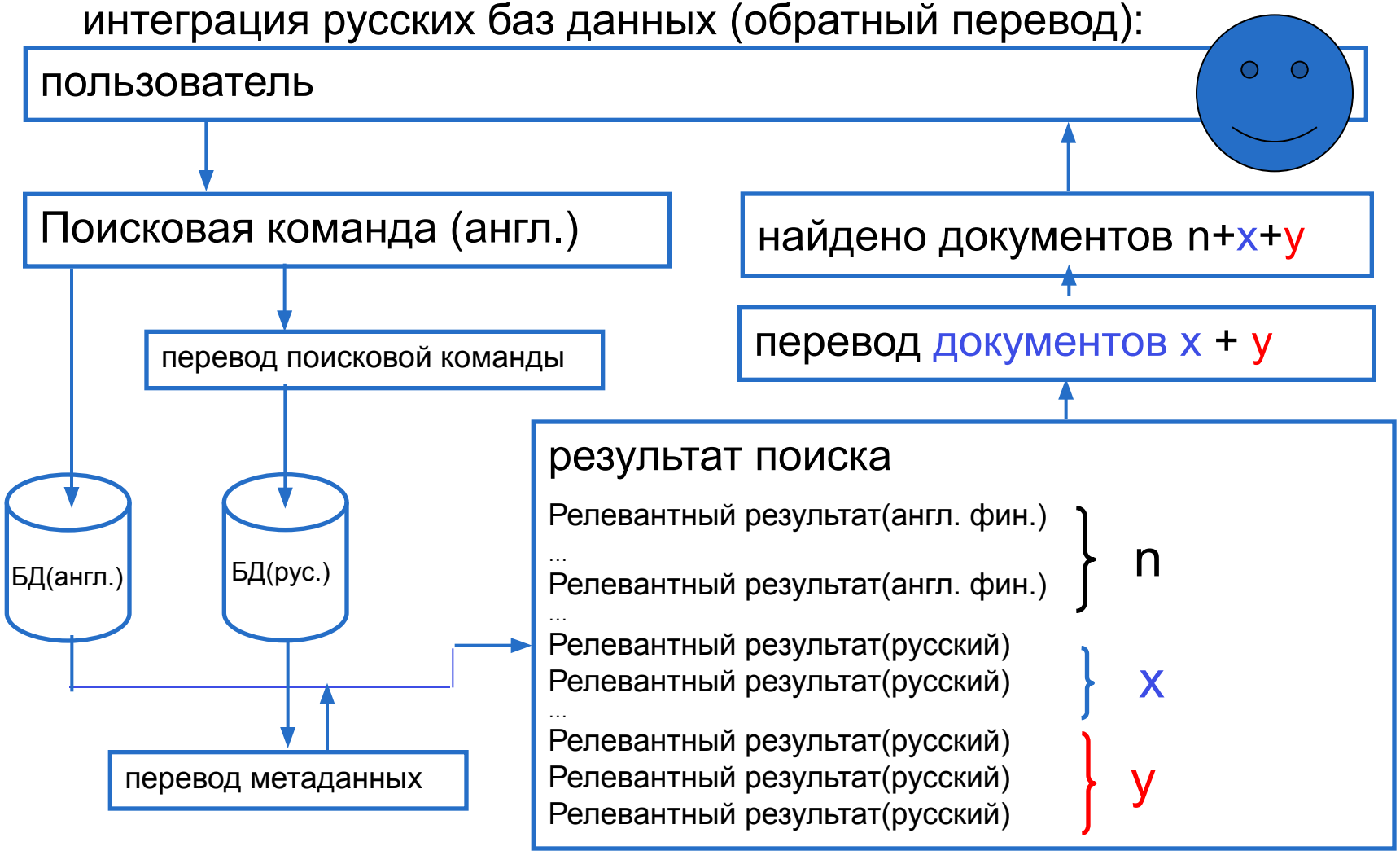
Оригинальное название	**Машинный перевод**	**Официальный перевод**
Федеральное агентство кадастра объектов недвижимости	Federal agency of a cadastre of objects of the real estate	Federal Agency of Real Estate Cadastre
Федеральная служба по надзору в сфере природопользования	Federal service on supervision in sphere of wildlife management	Federal Service for Ecology and Natural Resources Supervision
Федеральное агентство по строительству и жилищно-коммунальному хозяйству	Federal agency on construction and zhilishchnokommunalnomu facilities	Federal Agency for Construction, Housing and Communal Services

- Схема поиска информации с подключением машинного перевода:



- Обратный перевод - с английского на русский - позволит направлять поиск в русские поисковые системы путем перевода поисковых команд (Cross Language Information Retrieval)
- Перевод метаданных полученных результатов позволит повысить не только коэффициент полноты, но и коэффициент точности системы

- Схема поиска информации с подключением машинного перевода; интеграция русских баз данных (обратный перевод):



- Планы на будущее:

  Создание двуязычных (многоязычных ?) тезаурусов для поиска информации

- Создание систем репрезентации поисковых команд и документов, не зависящих от входного/выходного языков для многоязычного поиска информации

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

PROMT
TRANSLATION SOFTWARE

СПАСИБО!

RUSSIAinfo, университет Хельсинки

Компания ПРОМТ, Санкт Петербург

Таня Пурсиайнен

Дарьяна Цугульская