

Курс “Компьютерная лингвистика” на основе библиотеки NLTK

Рыжов А. А., магистрант МАИ

Чернышов Л. Н., доцент Финуниверситета

Компьютерная лингвистика

Решаемые задачи:

- Автореферирование текстов
- Рубрицирование текстов
- Машинный перевод
- Автоматическое извлечение фактов из текста
- Определение тональности текстов
-

Примеры задач компьютерной лингвистики (1)

Июл 14 | Авг 14 | Сен 14 | Окт 14 | Ноя 14 | Дек 14 | Янв 15 | Фев 15 | Мар 15 | Апр 15 | Май 15 | Июнь 15 | Июл 15 | Авг 15 | Сен 15 | Окт 15 | Ноя 15 | Дек 15 | Янв 16

Найдено: 1000000, Показывать по: плиткой, Сортировать: по релевантности

Лаврова возмутили обвинения в коррупции из Белого Дома

23:39 Подробности  0  0

... отражает официальную позицию Вашингтона. По сообщениям российского МИДа, Лавров в разговоре с Керри **выразил возмущение** этими обвинениями, назвав их надуманными. Вина за целенаправленное нагнетание напряженности в двусторонних делах полностью ...

Рецензия на Sublevel Zero



23:37 Riotpixels.com  0  0



... очередным противником и процедурной генерации локаций, Sublevel Zero **чудовищно однообразна**. Осилев первые два этажа, вы, скорее всего, увидите всё, чем может похвастать

Федеральная трасса под Ростовом стала похожа на поверхность Луны

23:35 Большой Ростов  0  0



... : множество аварий и пробка. Таких дорог по Ростовской области немало. И возникает вопрос: то ли это **отвратительное качество** дорожно-строительных работ, то ли и вправду погода у нас такая, что никакие дороги не выдерживают ...

посоветуйте пожалуйста

23:37 7ya.ru  0  0

... купила но мелкой удобнее заниматься на своем инструментена электронном она говорит и звук не тот и **нажатие не нравится** хотя купила не дешевоеЛетом планирую звукоизолировать комнату но что делать сейчас с этим психом соседом ума ...

Html страницы как пример

23:37 Just Business- электронный деловой журнал  0  0

... Предоставляется Заказчиком Наполнение (контент) Совокупность информационного наполнения веб-сайта. Если **представленная концепция не удовлетворяет** требованиям Заказчика, последний предоставляет мотивированный отказ от принятия концепции с указанием ...

Владелец кафе «Кахети» в Лабитнанги рассказал о драке с участием прокурора



23:36 Губкинский портал  0  0

... прокуратуры города. Потасовка с оружием произошла накануне поздно вечером. СУ СК Ямала уже начало проверку. **поводом для ссоры стало недовольство** посетителей. Пришла компания гостей, в которой было пять

Примеры задач компьютерной лингвистики (2)

Тип источника:

Любой

Источник:

Не выбран

Программа:

Не выбрана

Рубрика

HR

PR

Авиа

Автомобили

Алкоголь

Безопасность

Власть

Военное дело

Главные новости

Гостиницы и рестораны

Другое

Здоровье

Игры

IT

Корпоративные

Культура

Логистика

Маркетинг

Машиностроение

Медиа

Медицина

Металлургия

Наука и техника

Недвижимость

НКО

Образование и наука

Общество

Общество и бизнес

Отдых и развлечения

Политика

Право

Пресс-релизы

Продукты питания

Происшествия

Промышленность

Работа/карьера

Региональные новости

Религия

Ритейл

Сельское хозяйство

Семья и дети

Спорт

Стиль жизни

Страхование

Строительство

Телеком

Транспорт

Туризм

ТЭК

Управление

Финансы

Химия

Экономика

Электроника

Собственность

Государственный

Независимый

Назначение

Агрегатор новостей/портал

Блог

Журнал/тематический источник

Информационное агентство

Новостное издание/газета

Официальный источник

Реклама/объявления

Служба новостей (ТВ-радио)

Социальная сеть

Охват

Национальный

Региональный

Страна:

Любая

Регион:

Любой

настройки по умолчанию

Преподавание лингвистики в ВУЗах

- Совершенно различающиеся списки тем
- Различные виды лабораторного практикума (зачастую его отсутствие)
- Зависимость качества учебного курса от конкретного преподавателя

Предлагаемый формат спецкурса для направления “Прикладная математика и информатика”

- Подготовленный список актуальных тем
- Библиотека NLTK
- Задачи к лабораторному практикуму
- Направления для научной работы

Библиотека NLTK

Преимущества:

- Открытость
- Признание
- Академическая направленность

Недостатки:

- Неполная поддержка русского языка

Теоретические вопросы курса

Практическая направленность курса обеспечивается примерами на основе работы NЛTK и других лингвистических систем.

Особенности:

- Набор актуальных тем
- Подкрепление теоретического материала демонстрациями на основе NЛTK

Лабораторный практикум

Лабораторные включают в себя 1-2 емких задач на весь семестр, которые захватывают большую часть тем курса:

- Определение тональности методами классификации (л.р. или д.з. 16 часов)
- Фильтрация спама на почте

Направления НИРС

- Различные направления для научной работы
- Близкое знакомство с open-source библиотекой NLTK и возможность поучаствовать в её доработке

Интернет-сайт курса

Адрес сайта:

<http://ucheba8fac.wix.com/linguistics>

Курс "Компьютерная лингвистика"

Главная

Основные вопросы

Задачи к курсу

НИР

NLTK

Обратная связь

На данном сайте вы сможете найти общий формат спецкурса по компьютерной лингвистике. В своей работе мы изучили и обобщили программы курсов различных университетов России и мира и хотим предложить вам общую учебную программу на основе библиотеки NLTK.

Дальнейшие направления работы

- Постановка курса “Введение в компьютерную лингвистику” на кафедре Вычислительной математики и программирования МАИ
- Отработка набора тем по компьютерной лингвистике
- Создание новых задач с использованием библиотеки NLTK