

разнообразии программ, предлагаемых вашими телефонами. У всех этих приложений очень много функций: тут тебе и счётчик калорий, и социальная сеть, и обработчик фотографий, и личный тренер. Но зачем же так много всего? Это лишь отнимает у вас ваше драгоценное время. А данное приложение предлагает простой, приятный глазу дизайн, интерфейс, в котором не сложно разобраться, и удобство в использовании.

Итак, что же включает в себя разрабатываемое приложение. Впервые загружая его на экран своего телефона, вы вводите своё имя и индивидуальные параметры. На основе введенных данных программа высчитывает:

- количество калорий, рекомендуемое для употребления в день,
- соотношение белков, жиров и углеводов (БЖУ), которое подходит лично вам,
- базовый обмен веществ (это количество энергии, которое вы расходуете каждый день без учёта активности).

В соответствии с полученными результатами вы теперь можете вести дневник вашего питания и движения. В самой программе есть база данных продуктов с данными об их калорийности и БЖУ, а также база данных с основными активностями и видами спорта. Эти базы данных пополняемы, то есть в любой момент времени пользователь может добавить тот или иной продукт или блюдо, вид спорта и т.д. Это позволяет сделать приложение сугубо индивидуальным, что является безусловным плюсом. Дальше все просто: каждый день вы добавляете продукты питания, потребляемые вами в течение дня, а также активности. Приложение показывает вам, сколько калорий вам нужно еще употребить. Тут все зависит от вашей цели. Если вы хотите похудеть, вам следует употреблять меньше, чем расходовать, если просто поддерживать форму, то количество затрачиваемых и потребляемых калорий должно быть одинаково, ну и соответственно, если массу наоборот нужно набрать нужно употреблять больше калорий. В этом приложение вам и поможет. Вы будете видеть, сколько вы съели, сколько израсходовали. Причем есть возможность отдельно отслеживать, сколько вы потратили за счет сделанных шагов, сколько за счет занятий, сколько просто за счет жизнедеятельности своего организма. А с помощью удобного календаря можно видеть данные прошедших дней, а также корректировать их по необходимости.

Разработанное мобильное приложение для отслеживания питания и активности – это максимально удобное средство для того, чтобы следить за своим образом жизни, которое всегда под рукой и не будет отнимать у вас много времени.

Исследование поддержано проектом CERES. Centers of Excellence for young REsearchers (Reg.no. 544137-TEMPUS-1-2013-SK-JPHES),



Список использованных источников:

1. Developing iOS 10 Apps with Swift by Stanford. [Электронный ресурс] – Режим доступа: <https://itunes.apple.com/ru/course/developing-ios-10-apps-with-swift/id1198467120?l=en> . – Дата доступа: 03.04.2017.
2. Apple Developer. [Электронный ресурс] – Режим доступа: <https://developer.apple.com/> . – Дата доступа: 03.04.2017.

## АВТОМАТИЧЕСКОЕ РАСПОЗНАВАНИЕ КЛЮЧЕВЫХ СЛОВ В ТЕКСТОВЫХ ДОКУМЕНТАХ

*Белорусский государственный университет информатики и радиоэлектроники  
г. Минск, Республика Беларусь*

*Бредихин Ю. А.*

*Калугина М.А. – канд. физ.-мат. наук, доцент*

В докладе описан подход к решению проблемы автоматического распознавания и выделения ключевых слов в текстовых документах, являющейся частью задачи поиска релевантных текстовых документов в сети Интернет.

Определение ключевых слов документа является актуальной задачей для различных систем обработки текстовой информации. Как и аннотация, план или конспект, набор ключевых слов представляет собой документ с меньшей детализацией, но лишённый при этом синтаксической структуры. Поскольку ключевые фразы отражают основную идею документа, от их извлечения зависит эффективность приложений по обработке естественных языков (Natural Language Processing, NLP): информационный поиск (Information Retrieval, IR), вопросно-ответные системы (Question-Answering, QA), автоматическое реферирование (Summarization). Таким образом, качество алгоритма распознавания и извлечения ключевых фраз оказывает непосредственное влияние на работу этих приложений.

Для определения набора ключевых слов документа обычно используют лингвистические и математические методы.

Лингвистические методы основаны на семантическом анализе слов и дают большую точность и полноту поиска, однако приводят к весьма ресурсоемким процессам, занимающим большое количество времени в случае обработки больших объемов текстовых данных [1].

Статистические методы основываются на численных данных о встречаемости слова в тексте. Это значит, что они основаны на частотных характеристиках слов, то есть за ключевые слова принимаются те слова и фразы, которые имеют наибольшую частоту появления в тексте. При первом же анализе метода возникает следующая проблема: наибольшую частоту имеют неинформативные слова, например, союзы и предлоги, которые не являются ключевыми [2]. На следующем этапе приходится столкнуться с тем фактом, что разные формы одного и того же слова интерпретируются как различные слова. Все это приводит к сложности точного определения частоты появления слова в тексте и, следовательно, к правильному выделению ключевых слов.

В данной работе представлены результаты разработки алгоритма распознавания ключевых слов на основе лексико-грамматического подхода. С одной стороны, он базируется на статистических методах, с другой стороны, и это является его отличительной чертой, при анализе текста им учитываются лексико-грамматические характеристики слов, например, части речи, правила словообразования и т.д. Это приводит к возможности приведения к каноническим формам, а значит, к более точному выделению необходимых ключевых слов и фраз.

При использовании данного подхода обработка текста происходит поэтапно.

1. Лемматизация текста, или автоматическое определение канонической формы слова: каждое слово в тексте приводится в начальную форму.

2. Тегирование текста, или автоматическое определение принадлежности слова к той или иной части речи: каждому слову в соответствие ставится тег, то есть класс слова, описывающий часть речи и грамматическую категорию слова.

3. Фильтрация текста по информативности тега: из текста исключаются слова с неинформативными тегами.

4. Частотный анализ: использование алгоритмов, основанных на статических методах выделения ключевых слов, однако лишенных приведенных выше недостатков.

Для подсчета частотности слов применяется модель TF-IDF.

Вводится инверсная частота, с которой некоторое слово встречается в документах коллекции, IDF (inverse document frequency) – это логарифм отношения числа всех документов ( $|D|$ ) к числу документов, содержащих некоторое слово  $t_i$ , (когда  $n_i \neq 0$ ):

$$idf(t, D) = \log \frac{|D|}{|(d_i \ni t_i)|}$$

А также вводится параметр TF (Term Frequency) - частота термина в анализируемом документе, отношение числа вхождения термина к общему количеству терминов в документе:

$$tf(t, d) = \frac{n_i}{\sum_k n_k}$$

Ключевыми в данном случае будут являться слова, набравшие наибольший вес. Слова с малым весом вообще можно не учитывать при классификации.

Было разработано приложение на основе описанного алгоритма. Данное приложение позволило решить поставленную задачу с точностью 0.86. Данный показатель является весьма высоким по сравнению со статическими методами выделения ключевых слов. Время работы снизилось в несколько раз по сравнению с алгоритмами, опирающимися на лингвистические методы. Это позволяет считать решение поставленной задачи успешным.

Список использованных источников:

1. Гринева, М. Анализ текстовых документов для извлечения тематически сгруппированных ключевых терминов /М. Гринева, М. Гринев //Труды Института системного программирования РАН. Т.16. – 2009. – С. 155-165
2. Выделение ключевых слов в русскоязычных текстах / Ершов Ю.С // Молодежный научно-технический вестник [Электронный ресурс]. – Режим доступа : <http://www.sntbul.bmstu.ru/file/out/730754>. – Дата доступа: 20.03.2017