**Разработка алгоритма семантической декомпозиции текста на ключевые элементы для решения задачи категоризации**

Содержание

Введение

1. Обзор предметной области

1.1. Классификация и категоризация текстов на естественном языке

1.2. Анализ основных методов (подходов) обработки текстов на естественном языке

Параграфом раньше Вы декларируете, что Вашей основной задачей магистерской диссертации является классификация и категоризация текстов на естественном языке и подчеркивается, что речь идет о текстах на естественном языке. Это особенность Вашего проекта и относительно новая, малоизвестная тема в такого рода исследованиях. Поэтому в первой обзорно-аналитической главе Вы должны дать информацию о том, что это такое и как это можно использовать в Вашей работе. Область, действительно очень широкая и все сюда тащить не надо. Было бы хорошо на основе изученной Вами литературы сузить изложение материала до алгоритмов, которые напрямую или косвенно можно применить для решения задачи категоризации, может быть даже для некоторых вспомогательных задач, связанных с категоризацией. Здесь будут не только алгоритмы, которые Вы будете разрабатывать далее, но и другие, которые в работе не будут использоваться. Желательно, чтобы этот обзор был так сказать достаточно богатым.

1.3. Семантическая декомпозиция как один из инструментов семантического анализа текстов

1.4. Сравнительный анализ существующих алгоритмов семантической декомпозиции и выбор базовых алгоритмов для реализации

Вы выбрали судя по всему 2 базовых алгоритма, которые будут программно реализованы: алгоритм семантической декомпозиции на n-граммы, алгоритм семантической декомпозиции на простые предложения. Хотелось бы, чтобы в рассмотрении участвовало бы больше алгоритмов и Вы показали бы, почему с Вашей точки зрения эти алгоритмы лучше. Я в принципе хотел бы, чтобы Вы нашли бы еще какой-нибудь третий алгоритм. Тогда Ваша магистерская диссертация только бы выиграла и приняла бы так сказать классический вид (Бог троицу любит) 😊. Может быть Вам рассмотреть возможность модификации алгоритмов на основе ключевых слов из Вашей бакалаврской работы. Или это просто разновидность алгоритмов на основе n-грамм? В любом случае я бы рекомендовал потратить некоторое время и постараться найти третий алгоритм для реализации.

1.5. Выводы

2. Постановка задачи

3. Описание теоретических основ разрабатываемых алгоритмов

Вот это как раз желательно оставить как есть. Это из требований к оформлению теоретической части работы. Я не знаю, выполнимы ли и какой объем наполняемости эти трех подпунктов пунктов 1 и 2, но мне кажется в Вашем случае есть и какая-то математика, связанная с этими алгоритмами и на основе которой формируются эти алгоритмы. Алгоритмы обязательно должны быть описаны формально до их программной реализации. Здесь могут быть элементарные блок схемы или диаграммы активности UML или еще какие-то нотации. Для описания сценариев можно использовать, например, BPMN-диаграммы.

3.1. Алгоритм семантической декомпозиции на n-граммы

3.1.1. Математические основы алгоритма

3.1.2. Графическое описание алгоритма

3.1.3. Сценарий применения алгоритма для решения задачи категоризации

3.2. Алгоритм семантической декомпозиции на предложения

3.2.1. Математические основы алгоритма

3.2.2. Графическое описание алгоритма

3.2.3. Сценарий применения алгоритма для решения задачи категоризации

3.3. Выводы

4. Программная реализация алгоритмов

4.1. Алгоритм семантической декомпозиции на n-граммы

4.1.1. Технические подробности программной реализации алгоритма  
4.1.2. Тестирование алгоритма

Нет, тестирование здесь это просто проверка программной реализации алгоритма на отсутствие ошибок и проверка его работоспособности, а исследование предполагает определение его отличительных потребительских свойств в предположении, что Вы работаете с полностью работоспособной проверенной Вами в результате тестирования программой.

4.1.3. Сценарий использования

4.1.4. Применение алгоритма семантической декомпозиции на n-граммы для решения задачи категоризации текста

программный пример его привязки к решению задачи категоризации. В этом же наша главная цель...

4.2. Алгоритм семантической декомпозиции на предложения

4.2.1. Технические подробности программной реализации алгоритма  
4.2.2. Тестирование алгоритма  
4.2.3. Сценарий использования

4.2.4. Применение алгоритма семантической декомпозиции на предложения для решения задачи категоризации текста

4.3. Выводы

5. Исследование

5.1. Исследование времени работы

5.1.1. Сценарий эксперимента

5.1.2. Методика эксперимента

5.1.3. Используемые программные средства

5.1.4. Результат эксперимента

5.1.5. Сравнение и оценка результатов

5.2. Исследование точности результатов

5.2.1. Сценарий эксперимента

5.2.2. Входные данные

5.2.3. Результат работы алгоритмов

5.2.4. Экспертная оценка

5.2.5. Сравнение с экспертной оценкой

5.3. Выводы

6. Коммерциализация результатов исследований в рамках НИР

Да, это допраздел, который был в последние годы в магистерских диссертациях. Он только один. Возможно  у него будет несколько другое название. Но у Вас будет консультант по этому разделу, он Вам все скажет.

Заключение

Список использованных источников

Приложения

Вам надо выбрать 2 - 3 алгоритма для программной реализации. Они могут быть в идеале новыми, они могут быть хорошо известными, но Вы обоснуете их выбор для решения Вашей главной задачи - категоризации текста и Ваш результат здесь их программная реализация, они могут быть известными, но модернизированными Вами.