

федеральное государственное автономное образовательное учреждение
высшего образования «Санкт-Петербургский национальный
исследовательский университет информационных технологий, механики и
оптики»

Факультет _____ информационных технологий и программирования
Направление (специальность) _____ Прикладная математика и информатика
Квалификация (степень) _____ Магистр прикладной математики и информатики
Кафедра _____ компьютерных технологий _____ Группа M4238

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

на тему

Выделение групп пользователей в социальных медиа по их
интересам и поведению на основе множества источников данных

Автор магистерской диссертации	Дмитриев С.С.	_____
Научный руководитель	Фильченков А.А.	_____
Руководитель магистерской программы	Васильев В.Н.	_____

К защите допустить

Заведующий кафедрой	Васильев В.Н.	_____
	«_____» _____	2016 г.

Санкт-Петербург, 2016 г.

Магистерская диссертация выполнена с оценкой _____

Дата защиты «_____» _____ 2016 г.

Секретарь ГАК _____

Листов хранения _____

Чертежей хранения _____

федеральное государственное автономное образовательное учреждение высшего образования
«Санкт-Петербургский национальный исследовательский университет информационных
технологий, механики и оптики»

АННОТАЦИЯ ПО МАГИСТЕРСКОЙ ДИССЕРТАЦИИ

Студент _____ Дмитриев С.С.
Факультет _____ информационных технологий и программирования
Кафедра _____ компьютерных технологий _____ Группа _____ М4238
Направление подготовки _____ Прикладная математика и информатика
Квалификация (степень) _____ Магистр прикладной математики и информатики
Специальное звание _____
Наименование темы _____ Выделение групп пользователей в социальных медиа по их интересам и
поведению на основе множества источников данных
Научный руководитель _____ Фильченков А.А., кандидат. техн. наук, доцент
Консультант _____

КРАТКОЕ СОДЕРЖАНИЕ МАГИСТЕРСКОЙ ДИССЕРТАЦИИ И ОСНОВНЫЕ ВЫВОДЫ

объем _____ 14 _____ стр., графический материал _____ – _____ стр., библиография _____ 0 _____ наим.

Направление и задача исследований

Целью данного исследования является создание алгоритма выделения групп пользователей социальных сетей на основе их социальных связей и поведения в социальных сетях

Проектная или исследовательская часть (с указанием основных методов исследований, расчетов и результатов)

В рамках данной работы предложен подход, позволяющий выделять подгруппы у выбранной группы пользователей в социальных сетях, основывающийся на социальных связях и видимом поведении на публичных страницах. В основе предложенного подхода лежат несколько методов и концепций: представление социальных связей в виде графа, случайные марковские поля, а так же семантический анализ. В качестве примера использования подхода взята группа футбольных болельщиков, и подгруппа радикальных футбольных болельщиков, а так же группа феменисток и подгруппа радикальных феменисток. Были использованы данные пользователей из социальной сети Vk.com. Достигнуты следующие показатели для группы футбольных болельщиков: ; для группы феменисток: . Данный подход нов и так же может применяться для других групп и подгрупп пользователей.

Экономическая часть (какие использованы методики, экономическая эффективность результатов)

Данная работа не предполагает извлечения экономической выгоды из полученных результатов

Новизна полученных результатов

В рамках описываемого исследования представлен подход позволяющий определять принадлежность пользователя к определенной группе на основе его социальных связей и публичного поведения в социальной сети. Полученный подход является способом построения модели, не применявшимся для решения подобной задачи ранее.

Является ли работа продолжением курсовых проектов (работ), есть ли публикации

Работа не является продолжением курсовых проектов. На тему диссертации имеются публикации. //СПИСОК-2016//... потом дописать

Практическая ценность работы. Рекомендации по внедрению

Полученный алгоритм дает возможность определить является ли член выбранной группы так же членом её подмножества. Это может быть использовано правоохранительными органами, т.к. алгоритм позволяет выделить, например подгруппы, склонные к бандитизму, пользователей потенциально более способных на совершение незаконных действий, нежели среднестатистический пользователь. Так же алгоритм может быть использован для усовершенствования таргетированной рекламы, например для выделения подгруппы фанатов определенного бренда из группы его покупателей.

Выпускник _____

Научный руководитель _____

« ____ » _____ 2016 г.

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	5
1. Обзор предметной области	6
1.1. Задача восстановления характеристик пользователя.....	6
1.2. Задача разделения пользователей на группы.....	6
1.3. Обзор существующих решений.....	6
1.4. Постановка задачи настоящего исследования.....	6
1.5. Выводы по главе 1.....	6
2. Описание исследуемого подхода.....	7
2.1. Общая схема решения	7
2.2. Граф социальных связей.....	7
2.3. Определение тематики публичных сообщений	7
2.4. Случайные марковские поля.....	7
2.5. Модифицированные случайные марковские поля.....	7
2.6. Выводы по главе 2.....	7
3. Реализация описываемого подхода.....	8
3.1. Сбор данных	8
3.2. Использование алгоритмов	8
3.3. Способы измерения качества результата.....	8
3.4. Результаты.....	8
3.5. Выводы по главе 3.....	8
4. Заключение	9
5. Список использованных источников.....	10
5.1. Рисунки.....	10
5.2. Листинги	10
6. Проверка сквозной нумерации.....	12
Выводы по главе 6	12
ЗАКЛЮЧЕНИЕ.....	13
ПРИЛОЖЕНИЕ А. Пример приложения	14

ВВЕДЕНИЕ

В данном разделе размещается введение.

ГЛАВА 1. ОБЗОР ПРЕДМЕТНОЙ ОБЛАСТИ

Пример ссылок на литературные источники: [example-english, example-russian].

1.1. Задача восстановления характеристик пользователя

1.2. Задача разделения пользователей на группы

1.3. Обзор существующих решений

1.4. Постановка задачи настоящего исследования

1.5. Выводы по главе 1

ГЛАВА 2. ОПИСАНИЕ ИССЛЕДУЕМОГО ПОДХОДА

2.1. Общая схема решения

2.2. Граф социальных связей

2.3. Определение тематики публичных сообщений

2.4. Случайные марковские поля

2.5. Модифицированные случайные марковские поля

2.6. Выводы по главе 2

ГЛАВА 3. РЕАЛИЗАЦИЯ ОПИСАВАЕМОГО ПОДХОДА

3.1. Сбор данных

3.2. Использование алгоритмов

3.3. Способы измерения качества результата

3.4. Результаты

3.5. Выводы по главе 3

ГЛАВА 4. ЗАКЛЮЧЕНИЕ

ГЛАВА 5. СПИСОК ИСПОЛЗОВАННЫХ ИСТОЧНИКОВ

В качестве примера таблицы приведена таблица 1.

Таблица 1 – Таблица умножения (фрагмент)

–	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
2	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
3	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51
4	4	8	12	16	20	24	28	32	36	40	44	48	52	56	60	64	68

Есть еще такое окружение `tabu`, его можно аккуратно растянуть на всю страницу. Приведем пример (таблица 2).

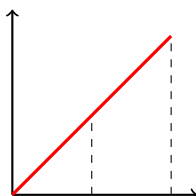
Таблица 2 – Таблица умножения с помощью `tabu` (фрагмент)

–	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
2	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
3	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51
4	4	8	12	16	20	24	28	32	36	40	44	48	52	56	60	64	68

5.1. Рисунки

Пример рисунка (с помощью `TikZ`) приведен на рисунке 1. Под `pdflatex` можно также использовать `*.jpg`, `*.png` и даже `*.pdf`, под `latex` можно использовать `Metapost`. Последний можно использовать и под `pdflatex`, для чего в стилевике продекларированы номера картинок от 1 до 20.

Рисунок 1 – Пример рисунка



5.2. Листинги

В работах студентов кафедры «Компьютерные технологии» часто встречаются различные листинги. Листинги бывают двух основных видов —

исходный код и псевдокод. Первый оформляется с помощью окружения `lstlisting` из пакета `listings`, который уже включается в стилевике и немного настроен. Пример Hello World на Java приведен на листинге 1.

Listing 1 – Пример исходного кода на Java

```
public class HelloWorld {  
    public static void main(String[] args) {  
        System.out.println("Hello, world!");  
    }  
}
```

Псевдокод можно оформлять с помощью разных пакетов. В данном стилевике включается пакет `algorithmicx`. Сам по себе он не генерирует флюатов, поэтому для них используется пакет `algorithm`. Пример их совместного использования приведен на листинге 2. Обратите внимание, что флюаты разные, а нумерация — общая!

Листинг 2 – Пример псевдокода

```
function ISPRIME( $N$ )  
    for  $t \leftarrow [2; \lfloor \sqrt{N} \rfloor]$  do  
        if  $N \bmod t = 0$  then  
            return FALSE  
        end if  
    end for  
    return TRUE  
end function
```

Наконец, листинги из `listings` тоже можно подвешивать с помощью `algorithm`, пример на листинге 3.

Листинг 3 – Исходный код и флюат `algorithm`

```
public class HelloWorld {  
    public static void main(String[] args) {  
        System.out.println("Hello, world!");  
    }  
}
```

ГЛАВА 6. ПРОВЕРКА СКВОЗНОЙ НУМЕРАЦИИ

Листинг 4 должен иметь номер 4.

Листинг 4 – Исходный код и флюат algorithm

```
public class HelloWorld {  
    public static void main(String[] args) {  
        System.out.println("Hello, world!");  
    }  
}
```

Рисунок 2 должен иметь номер 2.

Рисунок 2 – Пример рисунка

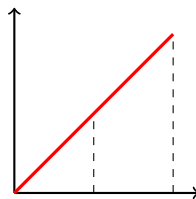


Таблица 3 должна иметь номер 3.

Таблица 3 – Таблица умножения с помощью tabu (фрагмент)

–	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
2	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
3	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51
4	4	8	12	16	20	24	28	32	36	40	44	48	52	56	60	64	68

Выводы по главе 6

В конце каждой главы желательно делать выводы. Вывод по данной главе — нумерация работает корректно, ура!

ЗАКЛЮЧЕНИЕ

В данном разделе размещается заключение.

ПРИЛОЖЕНИЕ А. ПРИМЕР ПРИЛОЖЕНИЯ