|  |  |  |
| --- | --- | --- |
| Автономная некоммерческая организация  высшего образования  «Университет Иннополис» | P:\Department Files\Business Development Department\Advertising Department\Public\Brand_material\Logo_PNG\Logo_02.png | Autonomous noncommercial organization  of higher education  “Innopolis University” |

**ДНЕВНИК СТАЖИРОВКИ (ПРАКТИКИ) ОБУЧАЮЩЕГОСЯ – STUDENT’S INTERNSHIP JOURNAL**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ФИО обучающегося – Student’s full name | | | | |  | Демиденко Андрей Аркадьевич – Andrey Demidenko | | | | | |  | |
|  | | | | |  |  | | | | | |  | |
| Номер студенческой группы – Student’s group No. | BS1-1 | | |  | | | | Номер студенческого билета – Student’s Record Book No. | 16B1110 | |  | | |
|  | | | | |  |  | | | | | |  | |
| Наименование направления подготовки – Field of study | | | | |  | Информатика и вычислительная техника – Computer Science | | | | | |  | |
|  | | | | |  |  | | | | | |  | |
| Уровень подготовки - Degree | | | | |  | Бакалавриат / Магистратура – Bachelor / Master’s degree  нужное подчеркнуть – underline as appropriate | | | | | |  | |
|  | | | | |  |  | | | | | |  | |
| Вид стажировки (практики) – Type of internship | | | | |  | заполняется сотрудниками университета – for employees of Innopolis University | | | | | |  | |
|  | | | | |  |  | | | | | |  | |
| Наименование стажировки (практики) – Name of internship | | | | |  | Административная / Научная / Бизнес ориентированная – Management / Research / Business  нужное подчеркнуть – underline as appropriate | | | | | |  | |
|  | | | | |  |  | | | | | |  | |
| Наименование организации – Name of organisation | | | | |  | Network Science and Information Technology Lab | | | | | |  | |
|  | | | | |  |  | | | | | |  | |
| Руководитель стажировки (практики) от Университета – Head of internship on behalf of Innopolis University | | | | |  | Джу Йонг Ли - Jooyoung Lee  ФИО, должность – Full name, capacity | | | | | |  | |
|  | | | | |  |  | | | | | |  | |
| Руководитель стажировки (практики) от индустрии\* - Head of internship on behalf of the company\* | | | | |  | ФИО, должность – Full name, capacity | | | | | |  | |
|  | | | | |  |  | | | | | |  | |
| Дата начала стажировки (практики) – Internship start date | | 13.04.2017 | | | | | Дата окончания стажировки (практики) – Internship end date | | | 25.05.2017 | |  | |
|  | | |  | | | |  | | | | | |  |

**Данные о стажировке (практике) – About internship:**

**Выполняемая задача - Task:**

|  |
| --- |
| Предсказание связей в социальной сети Вконтакте - Social Link Prediction on VK Network |

**Ход выполнения задачи - Procedure:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Дата начала – Start date** |  | **Дата окончания – End date** |  | **Выполняемый**  **этап задачи - Stage** |  | **Результат**  **выполняемого этапа задачи – Results of each stage** |
| **13.04** |  | **20.04** |  | Собрать тестовую выборку: 10 000 пользователей из Казани.  Collect test dataset: 10 000 users from Kazan. |  | Сделано.  Done. |
|  |  |  |  |  |  |  |
| **20.04** |  | **27.04** |  | Удалить 20% случайных рёбер. Сделать предсказание удалённых связей тремя методами: общих соседей, коэффициенты Адамик/Адара и Джакаарда.  Remove 20% random edges and predict them with three methods: Сommon Neighborhoods, Adamic/Adar and Jaccard coefficients. |  | Сделано  Done. |
|  |  |  |  |  |  |  |
| **27.04** |  | **04.05** |  | Рефакторинг написанного кода для оптимизации и удобного использования.  Refactor code for optimization and comfortable use |  | Сделано.  Done. |
|  |  |  |  |  |  |  |
| **04.05** |  | **11.05** |  | Определить какие параметры могут быть использованы для выявления промо-аккаунтов  Detect parametrs to filter promo accounts |  | Сделано.  Done. |
|  |  |  |  |  |  |  |
| **11.05** |  | **18.05** |  | Собрать базу со всеми активными пользователями сети.  Collect dataset with all valid users. |  | Сделано.  Done. |
|  |  |  |  |  |  |  |
| **18.05** |  | **25.05** |  | Получить для пользователей идентификатор города.  Get city ID for users. |  | В прцессе: ограничения API  In progress: API limit |

Информация верна – True and correct:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Обучающийся - Student |  | ФИО – Full name |  | подпись - signature |  | дата - data |
|  |  |  |  |  |  |  |
| Руководитель стажировки (практики) от Университета – Head of Internship on behalf of Innopolis University |  | ФИО – Full name |  | подпись - signature |  | дата - data |
|  |  |  |  |  |  |  |
| Руководитель стажировки (практики) от индустрии – Head of Internship on behalf of the company |  | ФИО – Full name |  | подпись - signature |  | дата - data |

**Отчет обучающегося по результатам прохождения стажировки (практики) – Internship Report**

|  |
| --- |
| Краткое описание достигнутого результата – A brief description of achieved results  Для исследования в качестве основных инструментов были выбраны язык Python и библиотека GraphTool (работа с графами).  Первая задача — собрать тестовую выборку из 10 000 пользователей по Казани. Вконтакте API предоставляет метод для поиска пользователей, но получить можно только первую тысячу. Для получения выборки был использован метод поиска по группе (группа города Казань), где можно использовать сдвиг (offset).  Далее для применения методов предсказания случайным образом было удалено 20% рёбер (связей), после чего с помощью метода общих соседей и коэффициентов Адамик/Адара и Джакаарда была сделана попытка предсказать удалённые рёбра. Результат оказался низким — наилучший показатель у предсказания с помощью коэффициента Адамик/Адара — 11.9% удачно предсказанных связей.  Для дальнейшей работы программный код был оптимизирован. Так же начата работа по оформлению модулей для независимого использования.  Возникает вопрос — почему в сети Вконтакте такой низкий уровень предсказаний? Было проведено небольшое исследование. Вконтакте, в отличие от Facebook, например, очень много промо-аккаунтов. Определить их можно по нескольким критериям: профиль, количество друзей, количество медиа-данных, активность. Но для того, чтобы проверить данные каждого пользователя нужно сделать не менее 5 запросов (для 1 миллиона пользователей — 5 миллионов запросов). API Вконтакте разрешает делать не более 3 запросов в секунду, при использовании метода execute — 75 запросов в секунду. Возникает большая проблема получения данных.  Для быстрого получения пользователей по городам было принято решение собрать локальную базу со всеми активными пользователями сети. Для этого было использован каталог пользователей и парсер. Далее необходимо для каждого пользователя получить идентификатор города, где возникает проблема ­с ограничением сети Вконтакте.  For research, Python language and the GraphTool library (working with graphs) were chosen as the main tools.  The first task is to collect a test dataset of 10,000 users in Kazan. VK API provides a method for finding users, but you can only get the first thousand. To obtain a dataset, a search method in groups (Kazan city group) was used, where you can use offset.  Then, 20% of the edges (bonds) were randomly removed for the application of prediction methods, after which an attempt was made to predict remote edges using the common neighbors method and the Adamic / Adar and Jakaard coefficients. The result was low - the best predictor with the Adamic / Adar coefficient was 11.9% of the successfully predicted edges.  For further work, the program code has been optimized. Also work has begun on the design of modules for independent use.  The question arises - why in the VK network such a low level of predictions? A small study was conducted. VK, unlike Facebook, for example, a lot of promotional accounts. You can determine them by several criteria: profile, number of friends, amount of media data, activity. But in order to verify the data of each user you need to make at least 5 requests (for 1 million users - 5 million requests). VK API allows you to do no more than 3 requests per second, when using the execute method - 75 requests per second. There is a big problem of getting data.  To quickly get users across cities, it was decided to build a local database with all the active users of the network. For this purpose, a user catalog and a parser were used. Next, it is necessary for each user to obtain a city identifier, where there is a problem with the restriction of the VK network. |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Обучающийся - Student |  | ФИО – Full name |  | подпись - signature |  | дата - data |

**Отзыв руководителя стажировки (практики) о работе обучающегося от Университета – Feedback from the Head of Internship on behalf of Innopolis University**

|  |  |  |
| --- | --- | --- |
| Качество выполненной работы и удовлетворенность полученным результатом – Quality of student’s work and satisfaction with achieved results |  |  |
|  |  |  |
| Сильные компетенции студента – Well-developed competencies of the student |  |  |
|  |  |  |
| Компетенции, нуждающиеся в развитии – Competencies to be developed |  |  |
|  |  |  |
| Общие рекомендации студенту – General recommendations |  |  |
|  |  |  |
| Результат аттестации по стажировке:  зачет/ не зачет – Internship results: passed/ failed |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Руководитель стажировки (практики) от университета – Head of Internship on behalf of Innopolis University |  | ФИО – Full name |  | подпись, печать – signature, seal |  | дата - date |

(Страница заполняется только для бизнес ориентированной стажировки, для иных стажировок страницу удалить – This page is for business internship only. Omit this page in case of other internship types.)

**Отзыв руководителя стажировки (практики) о работе обучающегося от индустрии – Feedback from the Head of Internship on behalf of the company**

|  |  |  |
| --- | --- | --- |
| Качество выполненной работы и удовлетворенность полученным результатом – Quality of student’s work and satisfaction with achieved results |  |  |
|  |  |  |
| Сильные компетенции студента – Well-developed competencies of the student |  |  |
|  |  |  |
| Компетенции, нуждающиеся в развитии – Competencies to be developed |  |  |
|  |  |  |
| Общие рекомендации студенту – General recommendations |  |  |
|  |  |  |
| Результат аттестации по стажировке:  зачет/ не зачет – Internship results: passed/ failed |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Руководитель стажировки (практики) от индустрии - Head of Internship on behalf of the company |  | ФИО – Full name |  | подпись, печать – signature, seal |  | дата - date |