

Про що цей документ

У цьому документі представлені результати діалогу зі службою Твіттер під час реєстрації на Порталі Розробника Твіттер. Мені двічі задавали уточнюючі питання. Але зрештою це виявилось корисним для мене і дозволило чіткіше сформулювати мету і методи проєкту.

Project explanation for Twitter API registration

Case# 0228523497 Twitter developer account application [ref:00DA0K0A8.5004w2Dpl3z:ref]

We would like to use the Twitter Data for assessment of air quality, particularly, by building smellscape or scentscape.

We plan to test the following technology chain:

1. Collecting the social media data (posts) by certain criteria, first of all, geotag (certain area, city), time period, key words presence, other
2. Analysis of the collected social media data (possibly using NLP) in order to get factor + emotion

We are going to use open Python libraries for collection, filtering and analysing the Twitter data, possibly with a use of NLP.

We certainly are going to use Tweets and, possibly, Likes. By analysing the text of Twitters we can determine air quality factor (PM, smell) and human attitude to this factor - positive, neutral, negative.

As a result we plan to obtain air quality factors connecting to geo location and period of time. So the results can be presented as dynamic maps or something similar

We foresee that the project result (air quality and its perception by citizens) can be interesting for the general public of, first of all, big cities or heavy industrial cities, and as such, for local governments of these cities (these might be municipal ecological departments or something similar).

The result of the project can be used by local governments of, first of all, big cities or heavy industrial cities as an indicative or alarm system to draw higher attention and undertaking timely countermeasures. The proposed system is an alternative to traditional instrumental methods of air quality monitoring and it has certain advantages prior to the latter ones. The primary advantage of the proposed approach is much

wider coverage by geographical distribution and by number of possibly evaluable physical factors without a need to deploy additional infrastructure.

Formidable requirement for clarification

A list of the government or public sector entities that will have access to Twitter content, or information derived from Twitter content, under this use case.

- Please note, we require an exhaustive and complete list of all public sector entities who would have access to Twitter content under this use case. Failure to provide an exhaustive and complete list constitutes a violation of the Developer Agreement, and may lead to enforcement action, including a potential suspension of access.
- This requirement does not apply to public sector entities with whom you may work or do business, but who would not have access to Twitter content, or information derived from Twitter content, under this use case.

Спробую розкрити свою відповідь. Мені здається, що виникло певне нерозуміння наших намірів. По-перше, **ми не плануємо виставляти вміст Твіттера будь-кому**. Натомість на базі аналізу відкритих Твіттів, ми плануємо визначати якість повітря у місті і чинники, що обумовлюють це. Отримані результати ми плануємо представляти у вигляді карти - smellscape. По-друге, нас дещо заплутує ваша вимога надати вичерпний перелік державних установ, що будуть мати доступ до вмісту Твіттера. З одного боку, до Твіттера мають доступ всі, тому вимога дещо дивує. З іншого боку, якщо йдеться про потенційних користувачів, то цей проєкт - це виключно наша ініціатива і наразі не має жодного зовнішнього фінансування, тому точно визначити **вичерпний** перелік користувачів на сьогодні не представляється коректним з огляду на відсутність критеріїв вичерпності. Тим не менш, відповідаючи на ваше запитання, ми уявляємо цільову групу наших користувачів таким чином: Це міська влада великих міст України та міст із заводами важкої індустрії в Україні. Також до цієї групи можуть належати українські НГО екологічного спрямування. До цієї групи можна зарахувати туристичний бізнес і ріелтерів в Україні.

I will try to reveal my answer. It seems to me that there is a certain misunderstanding of our intentions. First, **we do not plan to expose Twitter content to anyone**. Instead, based on the analysis of open Tweets, we plan to determine the air quality in the city and the factors that determine this quality. We plan to present the obtained results in the form of a map - smellscape. Secondly, we are somewhat confused by your request to provide an **exhaustive** list of government agencies that will have access to Twitter

content. On the one hand, Twitter is open and everyone has access to Twitter, so the requirement is somewhat surprising. On the other hand, when it comes to potential users, this project is exclusively our initiative and currently has no external funding, so to define a **exhaustive** list of users today does not look not correct given the **lack of exhaustiveness criteria**. However, answering your question, we see the target group of our users as follows: These are the city authorities of large cities of Ukraine and cities with factories of heavy industry in Ukraine. This group may also include Ukrainian environmental NGOs. This group includes the tourism business and realtors in Ukraine.

The specific use cases of your product or service by government or public sector entities, including information about:

- How you intend to analyze Tweets, Twitter users, or their content.

Наразі ми не маємо технічного завдання. Тим не менш наше загальне розуміння процесу аналізу таке: Ми плануємо аналізувати лише вміст твітів і не плануємо аналізувати користувачів Твіттера. За результатами аналізу вмісту твітів (можливо із застосуванням NLP) ми плануємо вибирати лише ті, що стосуються різноманітних прямих і опосередкованих оцінок якості повітря і супутніх запахів. Для цих вибраних твітів ми хочемо виявити фізичні/хімічні чинники, що формують цю якість повітря, геолокацію і час твіта.

We currently have no terms of reference. However, our general understanding of the analysis process is as follows: We plan to analyze only the content of tweets and do not plan to analyze Twitter users. Based on the analysis of tweets (possibly with a use of NLP), we plan to select only those that relate to a variety of direct and indirect assessments of air quality and associated smells, scents and odors. For these selected tweets, we want to identify the physical / chemical factors that shape this air quality, geolocation and tweet time.

- How Tweets and Twitter content will be displayed to users of your product or service, including whether Tweets and Twitter content will be displayed at row level, or aggregated.

На базі визначених фізичних/хімічних чинників, що формують якість повітря, геолокації і часу твіта ми плануємо представляти результати у вигляді карти - smellscape. Вміст твітів ми взагалі не плануємо представляти. Серед агрегованих даних має сенс представляти кількість вибраних твітів на певній просторовій ділянці, на базі яких була зроблена оцінка. Це потрібно для розуміння репрезентативності результату.

Based on determined physical / chemical factors that shape air quality, as well as geolocation and tweet time, we plan to present the results in the form of a map - smellscape. We do not plan to present the content of tweets at all. Among the aggregated data, it makes sense to represent the number of selected tweets in a particular spatial area, on the basis of which the assessment was made. This is necessary to understand the representativeness of the result.

- The core business purpose for your product or service.

Проблема інструментальних методів полягає в тому, що (1) практично неможливо забезпечити необхідну щільність датчиків, (2) датчики можуть вимірювати доволі обмежений перелік фізичних факторів, що визначають якість повітря, (3) розгортання додаткової інфраструктури та її підтримка вимагає значних коштів. Тому, підхід, що ми плануємо випробувати є альтернативою традиційним інструментальним методам моніторингу якості повітря і має певні переваги перед останніми. Основною перевагою запропонованого підходу є набагато ширше і щільніше охоплення географічним розподілом та кількістю можливих оцінюваних фізичних факторів без необхідності розгортання додаткової інфраструктури. Таким чином, ми розглядаємо наш проєкт як дослідження можливості створення альтернативної системи моніторингу якості повітря. Про комерціалізацію результатів можна буде говорити лише у разі отримання позитивних результатів цього дослідження.

The problem with instrumental methods is that (1) it is virtually impossible to provide the required sensor density, (2) the sensors can measure a rather limited list of physical factors that determine air quality, (3) the deployment of additional infrastructure and its maintenance requires significant resources. Therefore, the approach that we plan to try is an alternative to traditional instrumental methods of air quality monitoring and has certain advantages over the latter. The main advantage of the proposed approach is a much wider and denser coverage of the geographical distribution and the number of possible assessed physical factors without the need to deploy additional infrastructure. Thus, we consider our project as a feasibility study of an alternative air quality monitoring system. It will be possible to speak about commercialization of results only in case of receiving positive results of this research.