# Conversation Augmented Knowledge Extraction [CAKE] for SLACK

## Context:

Slack has become a very popular medium for team collaboration in today's digital world, but in case of large teams, it can easily become very chaotic. The EndCoronavirus.org slack workspace already has more than 300 public channels and around 4000 members. The number of messages and files shared each day is enormous (average of 3k messages per day). This scale of participation and richness of conversations can be mined by "team-talent-operations" to surface opportunities and connect dots. However, the manual effort is too much for a 100% volunteer project. We would like to offer our machine learning tool, code named CAKE, to help. It provides the functionality to easily and quickly identify topics and the users associated with each topic.

Since we don't have access to download all of the conversations in the ECV Slack workspace, we mocked up a prototype using data from a single channel: #analytics-scanning. This channel has around 1000 messages from 57 scanners, who research the internet to find and aggregate information for other ECV teams and projects. Here's a teaser question to illustrate how CAKE can help.

## How do we answer 'Which scanner should I contact for data on travel restrictions prevalent during covid19 ?' ?

### Before CAKE:

- ❑ Brainstorm and identify a comprehensive list of keywords that are expected in a post related to the topic in consideration (e.g.: 'travel', 'restrictions', countries', 'covid', etc.) Such keywords need to be determined for all the topics of importance, requiring a lot of manual effort.
- ❑ Reduce the subset of the messages to be examined out of the 1000 posts by filtering out the ones that don't have any of the identified keywords.
- ❑ Read through the large number of messages still remaining (57) to understand the language and semantics of each text to decipher the main topic, making it a more arduous task in case of a larger data space.
- ❑ One can always find the most frequently used words by each user, but this does not take the context of the message into consideration, and may not be useful in case of large number of users.

### After CAKE:

- ❑ Preprocessing the data with CAKE makes it very intuitive to cluster conversations and participants semantically
- ❑ The interactive real-time exploration reduced hours of slow and painful work into minutes of human-friendly exploration!

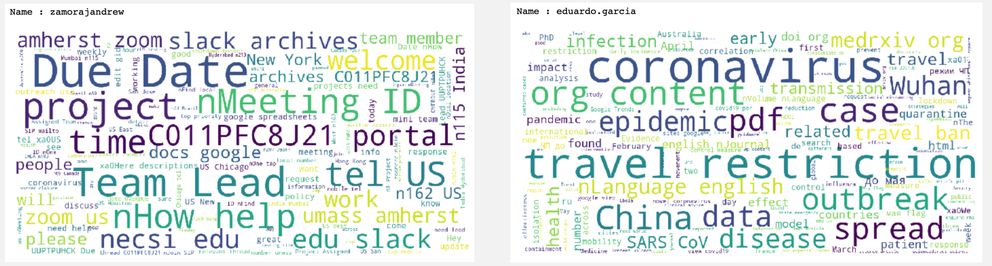


**Figure 1. Most frequently used words by Andrew and Eduardo**

| | Topic: Main Idea | Top Users |
|---|---|---|
| 0 | Admin Activities: Daily and Morning Updates | zamorajandrew, eduardo.garcia |
| 1 | Travel restrictions, quarantine policies around the world | eduardo.garcia,viljami.virolainen |
| 2 | Internet Scanning, Articles found | eduardo.garcia,viljami.virolainen,lapido,danielszev |

**Figure 2. identified users and their domain areas.**

## Recipe to bake this CAKE:

The application is developed by using cutting edge Machine Learning and NLP technologies from Columbia University and Metonymize [https://www.metonymize.com/]. It is a breakthrough in the team collaboration and team allocation processes.

Mapping this prototype to the entire Slack workspace should be a piece of CAKE! All we need is access to the logs and an initial set of Talent and Operations theme areas. Once proven to be helpful for day to day Talent and Operations team tasks, the capabilities can be easily extended to other areas. Further down the road, we could enable (i) live meeting transcript processing and build out (ii) a custom Slackbot for real-time assistance and engagement. Hopefully this tool allows the thousands of volunteers to find their projects and impact to EndCoronaVirus faster, together!