City Research Online

City, University of London Institutional Repository

Citation: Sathiyanarayanan, M. & Turkay, C. (2016). Determining and Visualising E-mail Subsets to Support E-discovery

This is the published version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/16154/

Link to published version:

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

City Research Online: http://openaccess.city.ac.uk/ publications@city.ac.uk

Determining and Visualising E-mail Subsets to Support E-discovery

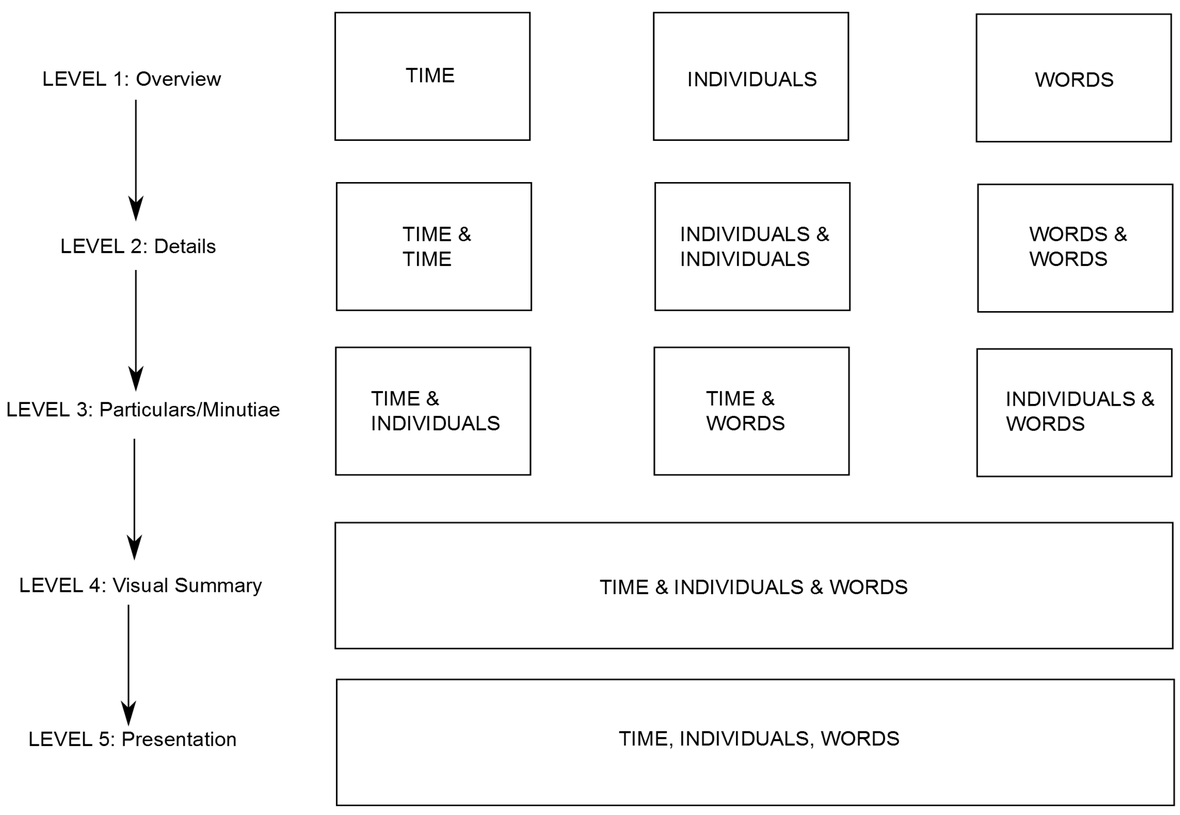
Mithileysh Sathiyanarayanan, Cagatay Turkay
giCentre, City, University of London

Introduction

Electronic discovery (E-discovery): Investigating emails collected over a period of time, manually, is a strenuous process and the current keyword search is very expensive.

So, there is a great need to determine, visualise and understand whether email subsets are normal or abnormal, pertinent or privileged, relevant (interesting) or immaterial in a quick time.

We proposed a multi-modal and multi-level approach which will generate automated visual representations using a manual search facility.



ONGOING WORK

- The prototype will be turned into a simple, powerful and analyst-friendly visualisation tool that will be tangible and feasible to use in E-discovery investigations.

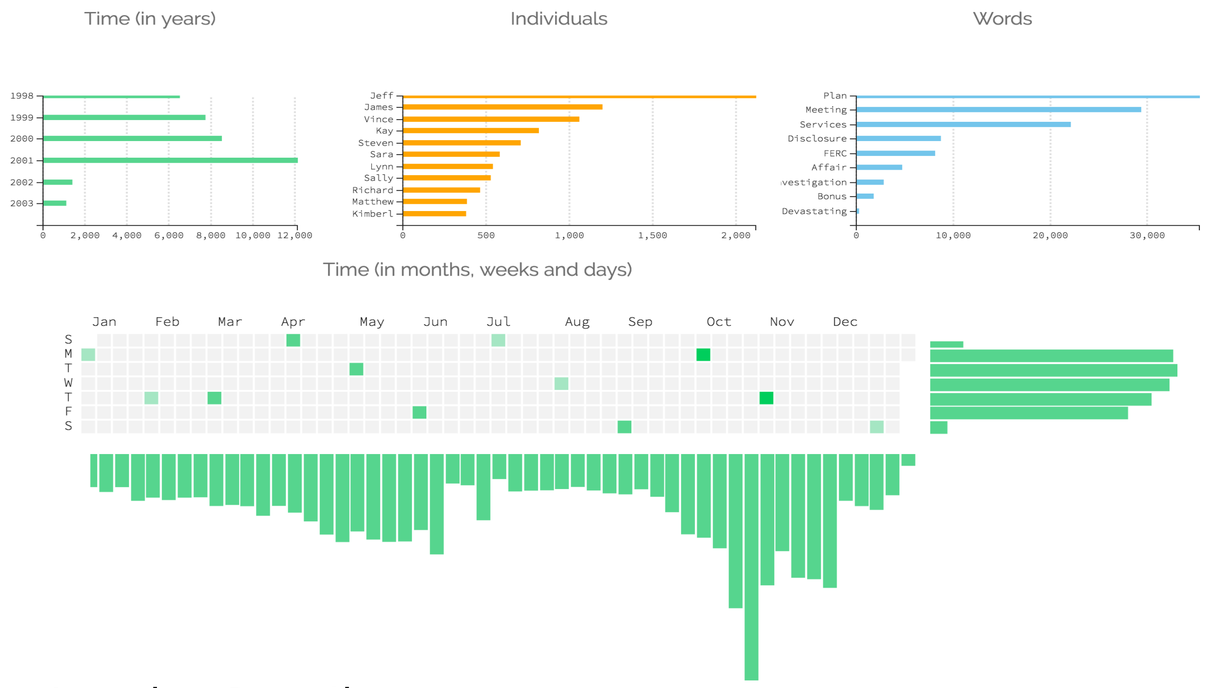
- Text analytics such as automated Named Entity Recognition or Classification of Email categories will be considered in order to provide valuable data preprocessing/analysis.

- Text visualization will be considered in order to provide effective views for the processed data.

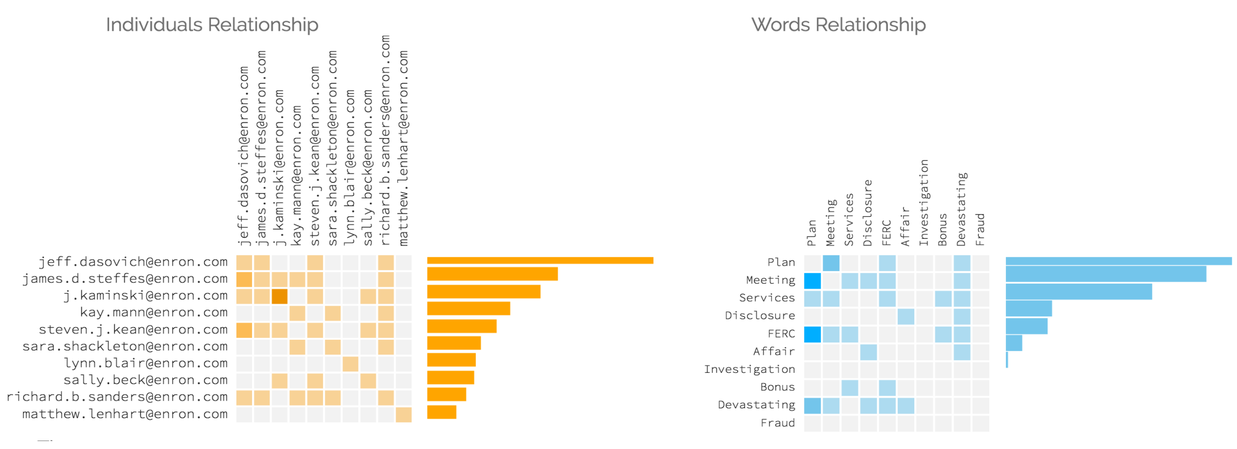
- The complete version of the tool will have user testing using Amazon Turk to evaluate the visualisation choices.

Design Approach

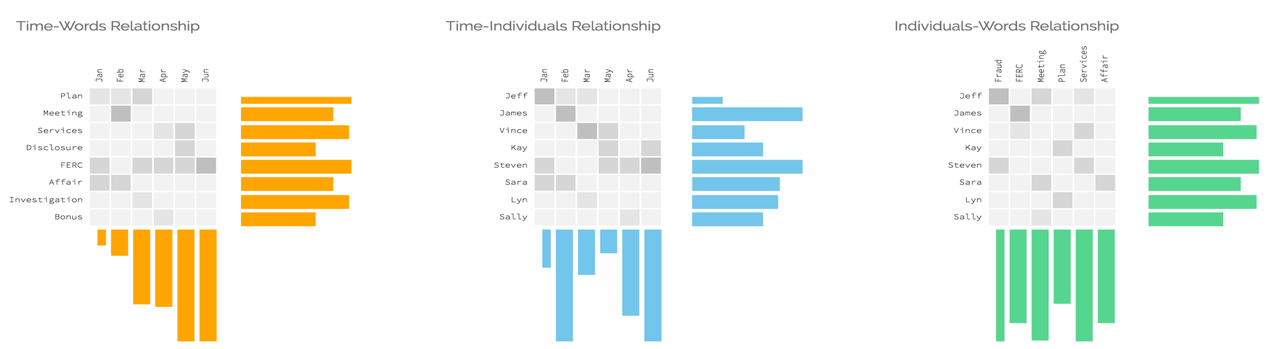
Level 1: Overview
Represents "anomaly" and "relevance" in data
Features:
- Choose time, individuals and keywords independently
- Data selection and filtering



Level 2: Details
Represents "interestigness" in data
Features:
- Choose time-time, individuals-individuals and keywords-keywords combination
- Data selection and filtering
- User interaction



Level 3: Particulars
Represents "Priviledgeness" in data
Features:
- Choose time-individuals, time-keywords and individuals-keywords combination
- Data selection and filtering
- User interaction



.

Discover insight from your data and automate workflows.
Redsift now in Public beta. Use QR Code for more details.

CITY UNIVERSITY OF LONDON EST 1894

RED SIFT
Data that speaks to you

giCentre