

## Contexte UIConsult

Le cabinet de conseil UIConsult souhaite se doter d'un outil d'aide à l'analyse des communications et des expertises à l'intérieur de ses équipes de collaborateurs. L'enjeu est pour les responsables de secteur de pouvoir produire un rapport d'analyse des échanges par emails des collaborateurs, appelé "RCom". Ces rapports sont réalisés à l'échelle individuelle et sont motivés par le besoin de suivre l'activité des collaborateurs, notamment pour la préparation des entretiens d'évaluation réguliers sur leurs performances.

La société est actuellement capable de récupérer la structure et le contenu des boîtes email des collaborateurs avec leur accord. Cependant en l'état les données sont difficiles à exploiter pour les responsables de secteur afin qu'il puisse préparer les RCom. Par ailleurs, il y a également pour enjeu d'extraire des emails et de façon la plus complète possible les contacts du collaborateur (email, nom, fonction, notamment) afin que ces informations de contact soient disponibles au format VCard (RFC6350).

### Exemple de données (<https://moodle.utt.fr/mod/resource/view.php?id=5389>) :

Message-ID: <12097696.1075857594249.JavaMail.evans@thyme>

Date: Wed, 13 Dec 2000 04:37:00 -0800 (PST)

From: john.arnold@enron.com

To: greg.whalley@enron.com

Subject:

Mime-Version: 1.0

Content-Type: text/plain; charset=us-ascii

Content-Transfer-Encoding: 7bit

X-From: John Arnold

X-To: Greg Whalley

X-cc:

X-bcc:

X-Folder: \John\_Arnold\_Dec2000\Notes Folders\sent mail

X-Origin: Arnold-J

X-FileName: Jarnold.nsf

remember when you said there is a reason they call them bear spreads?

bring up a chart of f/g or g/h.

f/g is tighter now than anytime since march 99 when ff1 was worth 2.50

amazing

Ce rapport devrait être synthétique et faciliter sa lecture avec l'utilisation de graphiques.

UIConsult font appel à vos équipes afin de réaliser un outil en ligne de commande qui permette aux chefs de secteur de gagner du temps dans la rédaction des RCom en extrayant les données importantes des boîtes emails et en préparant les graphiques au format SVG (et/ou PNG) à ajouter dans le rapport

Les responsables de secteur UIConsult ont remonté plusieurs indicateurs ou données d'intérêt dont ils ont besoin ou qu'ils souhaiteraient intégrer pour la rédaction des RCom :

- Nombre d'email échangés sur une période de temps (à défaut toute la boîte).
- Liste des "buzzy day" où un email a été écrit ou répondu le weekend ou entre 22h et 8h du matin le lendemain.
- Top10 des interlocuteurs pour un collaborateur donné
- Top10 des termes les plus utilisés dans les objets de messages pour une boîte mail donnée
- Visualisation en nuage de points (avec taille différente) du nombre d'échange entre collaborateurs pour une boîte mail donnée

Les responsables de secteur doivent pouvoir extraire des listes d'emails ainsi que lire et retrouver des emails selon différents critères de recherche (email, nom, prénom).

La liste n'est pas limitative et UIConsult est en demande d'autres suggestions d'indicateurs ou de vues pertinentes qu'ils pourront ajouter à leurs analyses.

UIConsult vous demande d'élaborer un logiciel en utilisant comme données de test deux extraits de sauvegarde de boîtes mail (extraite du corpus Enron Mail) qui correspondent aux données de boîte email de leurs employés. Les emails respectent le format standard RFC-5322. Ces données pourront être utilisées comme jeu d'essai d'exemple et de tests pour le projet logiciel demandé : <https://moodle.utt.fr/mod/resource/view.php?id=5389>

La solution proposée devra être réalisée en utilisant le langage Javascript et l'environnement Node.js en prévision d'une future intégration plus large de l'utilitaire sur le SI de la société.

## Bibliographie

IETF. Internet Message Format [En ligne]. RFC 5322, IETF, 2001, 51. [Consulté le 12 septembre 2018]. Disponible sur : <https://tools.ietf.org/html/rfc5322>

IETF. vCard Format Specification [En ligne]. RFC 6350, 2011. IETF, 2011, 74. [Consulté le 12 septembre 2016]. Disponible sur : <https://tools.ietf.org/html/rfc6350>

William W. Cohen. Enron Email Dataset [En ligne]. Version May 7, 2015 [Consulté le 12 septembre 2018]. Disponible sur : <https://www.cs.cmu.edu/~enron>

## Consigne

Un plan type vous est indiqué pour le cahier des charges. Chaque partie est accompagnée d'un volume indicatif. Bien entendu, vous ajouterez une page de garde avec un titre, une date, un numéro de version et les auteurs (le nom et la composition de votre équipe).

- Préface (1 page) : Expliquer brièvement à quels interlocuteurs est destiné le document (Audience), sa structure ainsi que les règles de révisions.
- Introduction (1 page) : Rappel général du contexte et des objectifs du projet logiciel.
- Glossaire (si nécessaire)
- Spécification générale des exigences (2-3 pages) : Courte introduction avec une liste des exigences fonctionnelles (10 maximum) et éventuellement non-fonctionnelles du logiciel à développer exprimée en langage naturel (comme vu en TD) avec un numéro de référence associé pour chacune.
- Spécification détaillée (8-15 pages):
  - Détail de chacune des exigences en format structuré (champs suggérés : Identifiant | Titre | Objectifs | Précondition(s) | Postcondition(s) | Entrées | Traitements | Sorties | Gestion des erreurs. Une description ou maquette devra être fournie quand une sortie de visualisation de données est attendue.
  - Une section « Formats de données » devra traiter particulièrement des différents formats exploités par l'application. Leur spécification sera précisée sous forme de Backus-Naur augmentée (ABNF).
  - Une section « Sémantique des données » où vous préciserez dans une spécification algébrique le principal type d'objet de l'application (Email ou Tweet), ses opérateurs d'ordre temporel, d'équivalence (entre même sorte). Vous donnerez également la spécification algébrique d'une collection permettant de contenir cette sorte, avec ses opérations et leurs axiomes associés.
- Conclusion (1 page) : Mise en perspective des objectifs du projet en regards des exigences formulées. Eventuels conseils pour l'implémentation
- Bibliographie (si nécessaire)
- Annexe (si nécessaire)

**Remise du livrable :** Dépôt sur Moodle du cahier des charges en pdf : <nom\_equipe>\_cc\_SujetA.pdf

Date de rendu : 11/11/20, 23h55

## **UIConsult**

The UIConsult consulting group needs to acquire a tool for the analysis of communication and expertise inside its consultant teams. The group is divided into several teams each led by a team leader. The team leader has to recurrently produce a report, called "RCom", that contains an analysis of her/his team members' email communication. These are individual level analysis that are motivated by the need to follow the employees' activities, especially for preparing the quarterly performance assessment interviews.

The company is currently able to access the structure and content of the email box of all employees (with their informed consent). However the data are difficult to use for the team leaders for the preparation of the RCom. Besides, UIConsult also would like to extract employees internal and external contacts information (email, name and function among other) in order to record them in vCard format (RFC6350) for further use in project. A good RCom has to provide an easy to read overview of one employee communication through the use of charts and key indicators.

**Exemple de données** (<https://moodle.utt.fr/mod/resource/view.php?id=5389>) :

Message-ID: <12097696.1075857594249.JavaMail.evans@thyme>

Date: Wed, 13 Dec 2000 04:37:00 -0800 (PST)

From: john.arnold@enron.com

To: greg.whalley@enron.com

Subject:

Mime-Version: 1.0

Content-Type: text/plain; charset=us-ascii

Content-Transfer-Encoding: 7bit

X-From: John Arnold

X-To: Greg Whalley

X-cc:

X-bcc:

X-Folder: \John\_Arnold\_Dec2000\Notes Folders\sent mail

X-Origin: Arnold-J

X-FileName: jarnold.nsf

remember when you said there is a reason they call them bear spreads?

bring up a chart of f/g or g/h.

f/g is tighter now than anytime since march 99 when ff1 was worth 2.50

amazing

UIConsult has called your teams in order to design and develop a command line tool that allows team leaders to save time and write better RCom. The tool should extract key data from an employee email box and to bake data charts in SVG format (and/or PNG) for inclusion in the report.

The UIConsult team leaders have fed back several relevant data that they need to access when they are writing the RCom report:

- Number of emails for a defined time period at day scale (by default all the email box)
- The list of "Buzzy days" where an email has been written or responded between 10:00pm and 8:00am of the next day.
- Top10 of an employee contacts
- Top10 of the most used words in an employee's emails subjects
- A scatter plot visualization (taking into account the amount of messages) of the number emails exchanged between internal and external collaborators for an email box

The team leaders should be able to extract email lists in order to be able to read and find emails according to different search criteria (email, first and last names).

The list of features is not limited and UIConsult is open to other ideas of relevant data to collect or process that they could use for their RCom.

UIConsult has shared with you a dataset of two mailboxes (corpus Enron Mail - <https://www.cs.cmu.edu/~enron>) that your teams could use as an example for the software project: <https://moodle.utt.fr/mod/resource/view.php?id=5389> .

The emails comply with standard format RFC-5322.

The software should be written in Javascript and be compatible with the Node.js environment in view of future integration with Synevent information system.

## References

IETF. Internet Message Format [En ligne]. RFC 5322, IETF, 2001, 51. [Consulté le 12 septembre 2018]. Disponible sur : <https://tools.ietf.org/html/rfc5322>

IETF. vCard Format Specification [En ligne]. RFC 6350, 2011. IETF, 2011, 74. [Consulté le 12 septembre 2016]. Disponible sur : <https://tools.ietf.org/html/rfc6350>

William W. Cohen. Enron Email Dataset [En ligne]. Version May 7, 2015 [Consulté le 12 septembre 2018]. Disponible sur : <https://www.cs.cmu.edu/~enron>

## ToDo

A standard outline for the software specification document is provided in what follows. A recommended length is proposed for each part. You will of course have to add a cover page with a title, a date, a versioning reference number as well as the name of your team and the list of its members.

- Preface (1 page) : Briefly explains the intended audience of the document, the document structure and associated revision process.
- Introduction (1 page) : States the context and aims of the software project.
- Glossary (when needed, especially for the acronyms used inside the document)
- Client level specification (2-3 pages) : Introduce the list of the software functional specifications (10 maximum) as well as possible non-functional specification for the software. Use natural language (as seen in TD) and provide a reference ID for each spec.
- Detailed specification (8-15 pages):
  - Detail of each specification using structured language formalism (suggested fields : Identifier | Title | Objective(s) | Pre-condition(s) | Post-condition(s) | Input | Process | Output | Error handling. Provide a description or mockup of the data visualisation output when needed.
  - A section "Data format" has to specifically presents the syntax of the data format that will be used by the software. The specification have to respect the Augmented Backus-Naur Form (ABNF).
  - A section "Data semantics" will use abstract data type formalism for the main object type used by the software (Email or Tweet) with its temporal order and equivalence operators (between a same type). You will also provide an abstract data type specification for a collection that contains such objects with its operations and associated axioms.

- Conclusion (1 page) : Highlight the relevance of the proposed specification regarding the project aims. Provide any useful advice for the development team.
- Reference (if needed)
- Appendix (if needed)

**Deadline for deliverable 1:** 11/11/20, 23h55

Upload your team specification document in pdf : <nom\_equipe>\_cc\_1.pdf