

Recommendation system for emails auto-completion using business process information

I- General Context :

The internship topic will be conducted in the context of a phd project consisting of mining business process knowledge from communication flows (e.g. emails, instant messaging) of a company. **Business processes (BP)** are used to manage or guide the interactions between a set of actors (for example by providing the sequence of activities to be executed) in order to achieve a business objective (e.g. recruitment, purchase of a quantity of product, ...).

BP knowledge are in general of three types :

- Actors' perspective knowledge (e.g. the set of employees that collaborate in order to execute an activity, the type of contribution made by them)
- Data Perspective knowledge (e.g.: business data (price, product quantity) that are useful when executing business process activities ...)
- Activity perspective knowledge (e.g. The sequence of activities necessary to achieve the objective of a given business process).

II- Goal :

The internship relies on exploiting the results obtained from the analysis of emails for discovering BP in order to develop a recommendation system (implemented in the form of a plugin for example). **For confidentiality reasons, the student will be led to exploit the results obtained from the analysis of a public database (Enron).**

The recommendation system aims to recommend for BP stakeholders the different kind of BP knowledge (Actor, data and activity) when typing or receiving their emails. The following functionalities are envisaged for the moment but the student can propose others :

1) Auto-completion of emails :

a) Recommendation of business data values to be entered:

Email users can type business data values when introducing an activity in an email. These business data values could (1) be previously introduced in other emails in similar contexts, (2) be previously stored in a different information system (e.g. ERP) intended for supporting BP execution when executing another activity in the context of the same BP, or/and (2) depend on previous typed values in the same emails. The goal is to consider the history of the typed business data values in emails or in other information systems to recommend the suitable values to be typed to emails writers.

Taking the example of the following email, the administrative assistant sends an email to notify a candidate (Richard) about organizing an interview. In such cases, she uses to mention the interviewers' names (e.g. Vince Kaminski, Stinson Gibner, Vasant Shanbhogue). These interviewers' names depend generally on some data mentioned in previous emails like job offer name or team/departement responsables (who proposed the job offer). When writing similar emails, the recommendation system should recommend such names to the administrative assistant by taking into consideration for example similar emails sent to other candidates or the information of the person name that has submitted the offer job in the recruitment information system .

Email :

Good morning Richard:

Your resume was forwarded to Vince Kaminski and the Research group and they would like to conduct a telephone interview with you at your convenience.

Please give me some dates and times that you would be available and I will coordinate the schedule. Also, the telephone number you wish to be contacted at.

The telephone interview will be last approximately an hour and the interviewers would be:

*Vince Kaminski Managing Director, Research
Stinson Gibner Vice President, Research
Vasant Shanbhogue Vice President, Research*

Thanks Richard and we look forward to hearing from you.

Regards,

*Shirley Crenshaw
Administrative Coordinator
Enron Research Group
713-853-5290
email: Shirley.Crenshaw@enron.com*

b) Recommendation of regular expressions used to express the activity (in the form of text templates to be filled for example): Some employees adopt regular writing style when introducing activities in their emails. Once detecting such behavior, the recommendation system must recommend to these employees regular expressions used by them to introduce their activities. This can be ensured in the form of text templates to be filled by business data values that vary from one email to another. Taking the example of the same email introduced previously, supposing that the administrative assistant uses the same textual content when contacting candidates and she changes only the value of three types of business data : (1) candidate name, (2) offer or department name (if it is present) and (3) interviewers names. In such case, the recommendation system must recommend to this assistant a text template to be filled by the values of these three business data types.

The recommendation system can recommend also simply some frequent expressions used to express the same activity or only the business data types that must be mentioned in the email once introducing an activity.

In general, the recommendation will depend on variant elements such as the typed email, the employee who is writing it, the type of the detected activity in the email or in a previous email, the identified profile of the employee ...

2) Recommendation of activity names related to one email: e.g. Once knowing the activity introduced in the received email, the recommendation system would recommend to the receiver the next activity that he must executed in response.

3) Recommendation of email recipients : e.g. Once identifying the activity introduced in an email before sending it, the recommendation system would recommend the recipients to the email writer (for example according to the actors that frequently collaborate with him for executing the same activity)

III- Missions:

The student could be based on some results of some works obtained after analyzing emails to discover some perspectives of business processes from emails [1,2]. These works discover: (1) activities and their business data from emails using a pattern discovery approach [1], and (2) actor perspective of activities using speech act detection from emails [2]. Results are obtained after carrying out experimentation on a public dataset Enron [3] and a part of them will be provided to the student (in term of activities and business data discovered from emails for example). Patterns that were discovered in the same context could be provided also to be used to detect activities from emails or to adapt them in the context of the recommendation system (e.g. by integrating some rephrasing techniques so that patterns will be understandable by emails' users).

The student will be invited then to :

- 1) Make a literature research concerning recommendation systems and the main existing algorithms.
- 2) Understand the context of business process discovery from emails in general (e.g. [4]) and the works [1,2] on their results he could rely on.
- 3) Propose solutions or adaptations of existing recommendation algorithms in our context
- 4) Propose adaptations of the results of previous works [1,2] to be exploitable in our context (or propose other independent solutions)
- 5) Implement the proposed solutions (in the form of a plugin for example to be installed in an emailing system). The implementation should ensure the different functionalities of the recommendation system in real time.

6) Conduct experimentation to validate the proposed recommendation system and propose simulation scenarios for the implemented solution.

References

- [1] Elleuch M., Ismaili O.A., Laga N., Gaaloul W., Benatallah B. (2020) Discovering Activities from Emails Based on Pattern Discovery Approach. In: Fahland D., Ghidini C., Becker J., Dumas M. (eds) Business Process Management Forum. BPM 2020. Lecture Notes in Business Information Processing, vol 392. Springer, Cham. https://doi.org/10.1007/978-3-030-58638-6_6
- [2] Elleuch Marwa, Alaoui Ismaili Oumaima, Laga Nassim, Nour Assy, Gaaloul Walid. Discovery of Activities' Actor Perspective from Emails based on Speech Acts Detection. International Conference on Process Mining (ICPM), 2020.
- [3] <https://www.kaggle.com/wcukierski/enron-email-dataset>
- [4] Elleuch Marwa, Laga Nassim, Alaoui Ismaili Oumaima, and Gaaloul Walid. Discovering business processes and activities from messaging systems: State-of-the art. In 29th IEEE International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), 2020.