

# 1 Annotation Guidelines

by Veronika Hintzen

## 1 Introduction

The planned email classifier is supposed to tag incoming emails for the extent of the required attention the recipient should give the email.

For annotation only the most recent part of the mail is relevant (older emails that are appended will need to be filtered...)

## 2 Explanation of classes with corresponding tag

The classes are sorted in descending hierarchy. there are eight classes of which only seven are suited for further processing. If more than one class is suitable for the email, the one higher in the hierarchy is to be chosen.

### **Reaction required - RR**

The sender requires a reaction of the recipient.

This can be a specific request for a reply or call as well as a question asked or any other request for the recipient to become active. Also questions worded as a declarative sentence such as 'I don't understand, why ...' that imply the need/wish for clarification.

### **Action required - AR**

The recipient of the email needs to do become active in any way. The email might contain a confirmation link or a task/work assignment.

### **Appointment/Deadline - AD**

The email informs/reminds the recipient of a (new) appointment or project deadline, the rescheduling of the same.

A specific date, time or time frame needs to be given.

### **Invitation - I**

The email is an invitation to a work-related event with optional participation, such as an orientation, workshop or party.

This also concerns rather private parties or dinners with several colleagues attending, making this a networking event.

### **Contains Information - CI**

The email contains work related information that in contrast to the AD class is not time-sensitive and that does not require a reaction.

E.g. FYI-mails, like information about the results of a meeting or an evaluation or a reminder about work related best practices or the availability of resources.

Also, when the email itself answers a question but without itself asking something from the recipient.

### **Private Mail - P**

The email is of a private nature and contains non-work-related information or conversation but one can assume that the recipient would want to read it.

### **Non-relevant - NR**

The email does not require any work related action at all. This is for example the case with junk mail, newsletters, any kind of automatically sent email. Also Mails only containing a 'thank you'-Note are considered non-relevant.

### **Unsure - U**

The email does not non-ambiguously fit into any of the preceding classes. The tagging of the email needs to be reviewed and possibly a new class needs to be created for future annotation. emails of this class for now are not suited for further processing.

## **3 Alphabetical list of tags with corresponding class**

1. AD - Appointment/Deadline
2. AR - Action Required
3. CI - Contains Information
4. I - Invitation
5. P - Private Mail
6. NR - Non-Relevant
7. RR - Reaction Required
8. U - Unsure

## **4 Problematic cases**

This section discusses difficult labelling decisions.

The announcement of a birthday cake in front of a colleagues cubicle was labelled as *Invitation* after some discussion.

It is not really about any work assignments but is time relevant and presents an opportunity to network/mingle with colleagues and is thus work related.

For obviously private mail, the new respective class was created, since they would not really fit in with any other class

Ambiguity of language and missing context information: e.g. 'please reply if I should change something': *if* in the sense of *whether* or used as a conditional.

A really satisfying solution has yet to be found. For now, the idea is to err on the safe side and choose the category that is upper in the hierarchy.

Invitations to events sent in automated newsletters - annotation depending on sender and content:

Is the sender a Enron-adress and/or the event very probably work related → *Invitation*

Does the sender and content look to be non work related → *Non-Relevant*

Mail from colleagues discussing private meet-ups:

Could it be a work dinner -> one of the work-related categories depending on other content

Does it sound more like a private date or meet-up → *Private*