14th ICCRTS
"C2 and Agility"

Achieving Agility by Improving Information Quality in Email Exchanges by Identifying Entities and Related Objects

Track 10, Collaborative Technologies for Network-Centric Operations

Dr. Marek Kowalkiewicz, Hans Peukert, David Lincourt

Point of Contact
Hans Peukert

Chief Strategist
Global Defense & Aerospace Sector
SAP AG
Dietmar-Hopp-Allee 16
69190 Walldorf
Germany
Telephone +49 6227 7 41296
E-mail Address: hans.peukert@sap.com

Abstract

E-mail exchanges are pervasive within defense organizations; rarely are they viewed as value-added by decision makers. Rather, it is a communication necessity that is not integrated with other aspects of decision making. It is then imperative that e-mail users be provided with higher contextual information relating to e-mails in order to improve decision making. For example, when sending e-mails, content is a static snapshot of the situation rather than the actual one when the receiver reads it – either for the first time or revisits it some time later. Similarly, when receiving e-mails, the reader has to manually jump to other systems to get the current situation unless the sender has provided that link in the e-mail.

We have developed in a proof of concept a tool that is integrated within the e-mail client that conducts text analysis and proposes potential contextual links to relevant systems for further analysis and decision making. The prototype has been developed with a focus on email exchange in military scenarios. Our approach creates an innovative user experience that is tailored to the specific information needs of the user. Specifically, the field of information extraction provides relevant tools to recognize objects that could potentially be enriched with additional information thus dramatically increasing the value of e-mails and similarly electronic documents.

Keywords: Text Analysis, e-mails, contextual links, Operationalizing the back office,

1 Introduction

Electronic mail is – without a doubt – the most pervasive asynchronous communication method. The speed of delivery and relative reliability of email exchanges across heterogeneous networks are making it a primary candidate for use in military situations where no proprietary network configuration can be assumed. In the peace time, when there is no need for measures that overcome numerous shortcomings of electronic mail, it is also a standard communication tool.

Apart from time, there are virtually no costs of creating and sending emails. On top of obvious problems created by that situation, such as overwhelming numbers of spam messages, it also means that email authors often do not try to make sure that the information in their messages is accurate or complete. There is no incentive for creating high quality, information rich (or at least complete) emails.

On top of military scenario email is a powerful means of communication for casual users. It is even more important for businesses. Email clients are often the only application that is running throughout the business day. Radicati estimates that in 2009 41% of each business day will be spent managing email [1]. Unsurprisingly, electronic messages are vital for conducting business. In business context, missing or misinterpreting just one email (for example containing sales leads) may lead to missing revenue. Military scenarios are no different. Therefore tools that streamline work with email attract a lot of attention from the business world and should also become more pervasive in the defense environments, where overlooking an email, or a piece of information in an email message can have far reaching consequences.

While the speed of delivery is one of the advantages of e-mail, it cannot be assumed that the actual intake, or analysis, of information included in an email takes place immediately after it has been sent. In other words, the content in an email is usually a snapshot of the situation at the time of creating the email, not at the time of reading it by the recipient. That in turn means that there is no certainty that the information contained in an email message is still up-to-date. In case of important messages, the recipients might be forced to verify the information, often by getting access to the source information (when possible) or contacting the sender using synchronous communication means.

Currently in defense organizations the key users (decision makers, information workers, business users) spend over 80% of their day in a business productivity suite, such as Microsoft Office or Open Office. Many key users only require sporadic access to enterprise applications (such as enterprise resource planning, or ERP, systems) for operational or corporate processes, and thus it would be preferable to combine the enterprise applications' functionality with business productivity suites. Only "power users", focusing on very specific areas of operations, are used to access operational and corporate processes in the ERP suites. All the other categories of users predominantly use business productivity suites, and one can assume that their abilities to use ERP systems are limited.

The Defense Forces are characterized by a number of major drivers in the area of empowering information workers:

- *A need for improved fact based decision-making in the enterprise.* The users in defense organizations rely heavily on ability to access heterogeneous sources of

information in order to make informed decisions. The higher the value of information sources chosen to support the decision making process, the higher the quality of decisions made.

- *A need for more productivity from the information workers.* Especially in high stress situations, number of messages exchanged may be overwhelming. In such cases tools and methods that increase the efficiency of information workers (for instance through providing relevant information) are invaluable.
- *Desire to increase participation in utilization of military functional systems, e.g. Enterprise Resource Planning (ERP) applications.* The ERP applications contain valuable data that can be used in decision making processes as well as streamline the work of information workers. The defense force users not having access to ERP data would have to contact the "power users" to get requested information. Ability to provide, even simplified, access to ERP data from within business productivity suites is very compelling. With such approach the back office information, contained in ERP suites can effectively be operationalized.
- *Limited motivation to "learn" military functional systems, e.g enterprise applications*. As in any other scenario, introducing new information systems may put additional burden on users – mostly through required additional training. Ability to embed enterprise application functionality in a familiar environment reduces the burden.

In many defense scenarios the situation develops quickly, and while the information contained in an email might be a valuable one, it often needs to be updated, and new facts need to be brought up. For instance, if an email states that a group of soldiers has been dispatched to an emergency area, it might be reasonable to bring up more information about the situation and – if needed – suggest dispatching additional troops. In other scenarios, the information contained in an electronic document might need to be verified. Or an email sender might be requesting information that resides in an internal system.

In all of the above scenarios, the recipient of the document needs to refer to external systems and access the information. From the process point of view, the recipient needs to be able to (1) identify their information need and then (2) identify the source of the information needed to fulfill the need, (3) access the source, and finally (4) perform actions that will fulfill the need. At least some of these steps (recognizing the need, identifying sources, accessing the sources) can be automated, and investigating possibilities of automating these steps was the main topic of our work in the past months, results of which are described in this document.

We have decided to revert to the domain of Natural Language Processing (NLP) in order to fulfill the abovementioned process. NLP has reached a state of maturity where commercial applications are possible and reliability of such applications becomes acceptable [2]. Although some of the areas of NLP, for instance Information Extraction (IE) or Named Entity Recognition (NER), are particularly advanced, there still is no ``killer application" that can demonstrate high potential of NLP [3]. That holds for both enterprise and non-enterprise worlds.

There are some examples of NER applied in popular applications. For instance Google and Apple have been using NER in their email applications to recognize such entities as dates and addresses (temporal and geographical entity recognition), to provide ability to create new events and view maps, respectively. These systems are not extensible by end-users in any way, and there is no possibility to enable recognition of new types of entities or provide other, alternative actions that can be performed with the recognized entities.

Based on this observation, we believe that, by applying Service Oriented Architecture (SOA) concepts in NER and IE domains, we can provide more flexibility to users. Ability to add new types of entity recognition on top of standard ones is especially important in enterprise context. Simple date, place name, or human name recognition can be extended with features such as product name (code), customer name or contract number recognition, providing much more automation to the enterprise world. Ability to extend list of actions that can be performed on recognized entities brings benefits to both enterprise and non-enterprise worlds.

In our work we have analyzed numerous NLP methods and tools and as a result, we have built a proof of concept prototype, showing how email exchanges can be enriched by automatic suggestions of relevant information that, among others, help composing replies.

Applying NLP methods is only useful if it leads to providing end users with information or actions that are relevant and helpful. Since relevant actions (as well as the ranking of these actions) vary based on various factors, high flexibility and ability to replace and rank system components become useful.

We believe that it is possible to demonstrate the strength of the mentioned NLP methods by integrating them in a single system, using SOA paradigms. In this demonstration we showcase Yowie, a system that recognizes entities in documents and provides end users with related information and actions. Yowie is following SOA guidelines, and as such is fully extensible and can make use of external services. In our demonstrator we chose to use SAP ERP system as an example, however the solution is not limited to any particular vendor. Similarly, there is no strong requirement that Microsoft Office be used, leaving full flexibility of choosing a business productivity suite.

2 A need for agility

Addressing the defense requirements in the context of information processing needs to follow five principles:

1. *The informational needs of decision makers need to be addressed.* We consider decision makers as ones that have to use numerous systems to access and collect information crucial in their tasks. Especially in these cases, ability to combine the information in a “one-stop-shop for decision making” is important.
2. *Understandable, contextually relevant, and actionable knowledge needs to be provided.* Pure data, presented out of its context can be of no use if not presented properly. It is also important that the link between the context (such as contents of an email) and the the presented information is clear to the end user.
3. *The current generation of decision makers is comfortable with information processing applications, understands the related risks, and uses them responsibly.* Therefore a requirement was to provide end users with a system that they are familiar with.
4. *The decision makers should be able to interact with the system in order to improve its functionality in a non-technical way.* Ability to modify provided functionality, improve the behavior of the system through mechanisms such as self-learning systems are an expected functionality.
5. Agility of applications should be achieved through end user actions (including composition) and not custom development of applications. In other words it is the behavior of the users (such as marking selected information as more relevant in a

specific context) should be shaping the future behavior of the system. At the same time, the interactions should be as natural to the end users as possible.

3 The prototype

Yowie is a system aiming at providing a link between business productivity software and sources of information and services such as the Internet, local desktop, and enterprise systems. The link that Yowie is providing is based on the assumption that certain fragments of documents processed (viewed, created, sent, received) in business productivity software contain enough information to link to data objects and services available in external systems automatically.

In Yowie, we assume that it is possible to use known entity recognition methods and based on that provide relevant services to end users. Current architecture of Yowie ensures a system that is independent from a particular implementation of business productivity software, as well as from particular implementations of external systems, due to SOA-based approach. Yowie is a system that can be fully deployed locally or make use of client-server architecture, depending on requirements. It consists of four main components: a set of *plugins* for various document processing applications, a *core* acting as a mediator between all other components, a set of *extractors*, and a set of *service wrappers*.

- *Plugins* are written specifically for individual document processing applications and integrate into them, so that relevant content can be processed and enriched by Yowie. Plugins are responsible for forwarding contents of documents to Yowie and, after Yowie has processed them, for providing the application users with relevant user interface containing related information and service links.
- *Core* provides a connection between plugins and the main Yowie functionality. Core receives document from Yowie plugins and passes it to extractors and further service wrappers. The results from service wrappers are passed back to relevant plugins by Core.
- *Extractors* process content of the documents and recognize entities included in the documents. Extractors can be standard (such as temporal or geographical entity extractors), they can also be tailored (invoice number extractors for instance).
- *Service wrappers* take one or more recognized entities and perform actions. Service wrappers are services themselves, but they can also provide access to other services that are external to Yowie. Service wrappers can be implemented as Plugins for retrieving information from other applications. In this case, they are also called *guest plugins* to distinguish them from the (Host) plugins mentioned above.

Yowie exhibits the following novel characteristics:

- *Integration of NER and IE services.* Yowie is a system that can integrate various NER and IE services in one application following the SOA paradigm. Previous approaches were either hard coded or did not integrate various services at all.
- *NER and IE: entry points for service consumption.* NER and IE can be used for automatic service selection. Therefore, Yowie is an example of an application that can consume services offered by a service broker and automatically choose ones that are relevant.
- *A potential for an SOA-enabled killer application.* By integrating NER, IE, and SOA in end-user productivity applications, Yowie, or the idea behind it, has the potential of

becoming popular quickly and thus being one of the more visible examples of using SOA in end-user applications.

In Figure 1 we show a sample behavior of Yowie, where several types of entities are recognized in an email and relevant information and actions is presented. One can see “Yowie People”, “Yowie Dates” and “Yowie Places” types of categories. In the “Yowie People” case, recognized entities are enriched with information such as an email address and phone number, and links to perform actions (call, send e-mail) are provided. Recognized dates give an ability to create appointments, and recognized locations show a map and local weather.

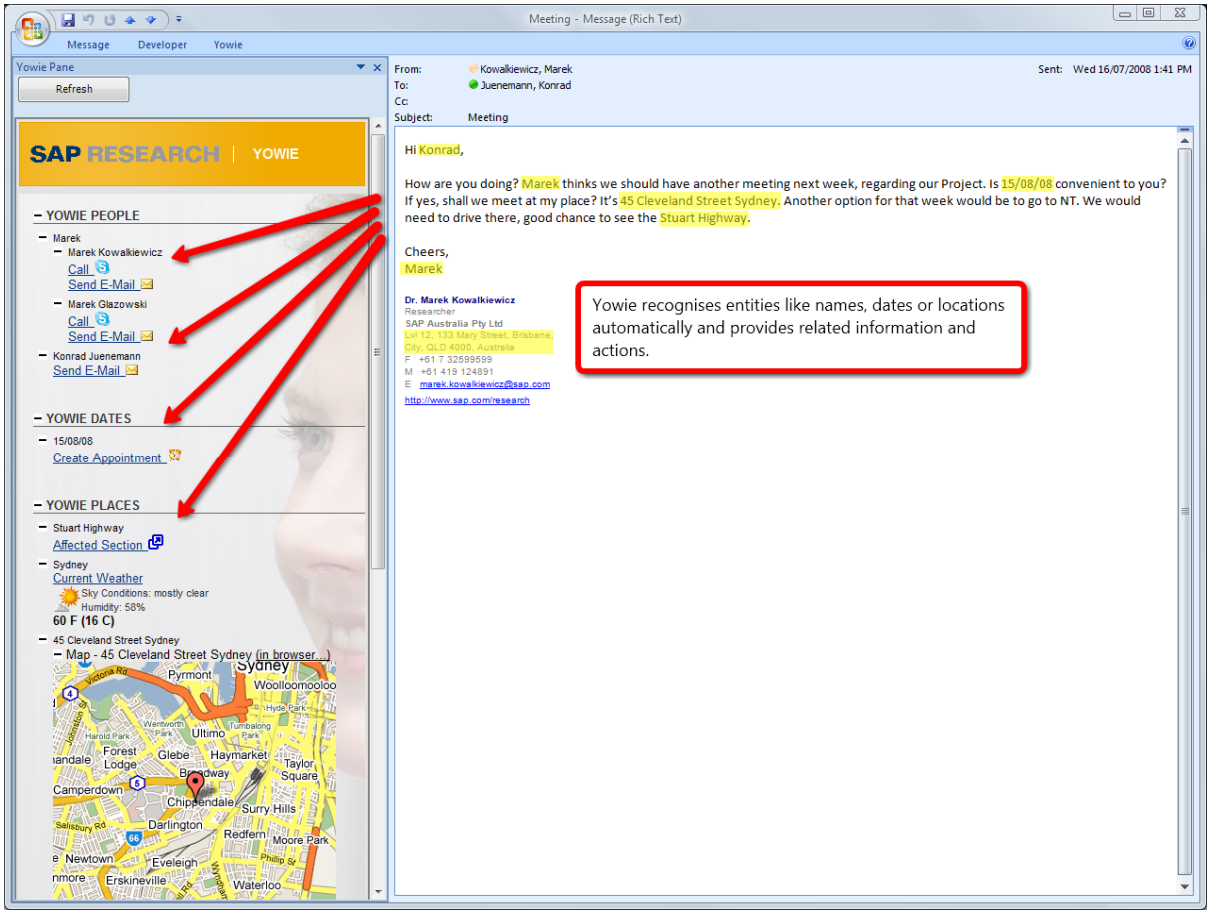


Figure 1 Analysis of email contents in a mail client embeded Yowie sidebar

The architecture of the system is fully modular, meaning that recognition of chosen types of entities can be turned on or off, and related information and actions can also be replaced. For instance, in a business scenario, when recognized a location (“Yowie Places”) a list of customers located within a certain distance from that location could be displayed.

4 Applicability of the prototype for military decision maker

One of the problems that military decision maker currently face is one of relating information out of an email with information stored in various heterogeneous systems in often structured form. As we observed, military users often “cut & paste” parts of emails’ content and process it in another application. They are also often forced to use other means of communication, such as a phone due to lacking knowledge or authorization to access a backoffice system.

In case of “cutting & pasting” the information the military decision maker is bridging the gap between the unstructured domain (such as email in our examples) and the structured domain of military backbone systems or the Internet. Actionable real-time knowledge is created only

through human actions, bridging the gap between the two classes of systems. While when accessing the structured domain the military decision maker is accessing real time information, in the unstructured email world, the email content remains static

With our prototype we are trying to understand whether such gap can be closed and automated. The prototype will automate these processes by identifying entities and related objects. Below (Table 1) we show a sample email message that could be exchanged in military scenarios and further show how our prototype can support collection of information in such a situation. In the email we have already highlighted entities that we believe are relevant to the recipient and which therefore should be enriched with additional information.

Table 1 A sample email in a defense scenarios wth highlited entities

Email from **Col Rob Moore** to Col Peukert Status **Operation Eagle Thunder** **Urgent: Blackhawk Tail 12345** of **131St of Army Aviation Brigade** earmarked for operation Eagle Thunder, Current position: Sydney, Operational Status INOPERATIVE Spare part **5230-12-1712346** for immediate repair missing. When will Spare Part be available in Sydney? Alternative Course of action: Is **202 Army Aviation Brigade** able to dispatch replacement Blackhawk with **ETA 28121700 Z** meeting point Bravo **Alice Springs?**

In its prototype implementation, Yowie processes the text of the email to identify, retrieve and display relevant additional information. The information is be displayed in a sidebar providing the military decision maker with real time information from other systems. In the prototype we show recognition of **Col Rob Moore** as an entity and display address book information together with his position in an organization chart. The information is actionable, so that a connection to Col Rob Moore can be made immediately. In this scenario a virtual meeting room can be opened. **Operation Eagle Thunder** is identified as a key word in Enterprise Search & Shows the latest Document on Eagle Thunder (Operation Plan). **Blackhawk 12345** is identified as weapon system (Master Equipment Location in the enterprise system used for prototyping) and the current Readiness Status is displayed via a customized Defense Widget showing Status red X (unable to fly). Further entities are also recognized and relevant information and services being presented to the information worker. Figure 2 below shows a snapshot of the user interface with Yowie providing relevant information as described above. Through these means the recipient of the email can automatically see if information received via the email is still relevant and not overcome by events.

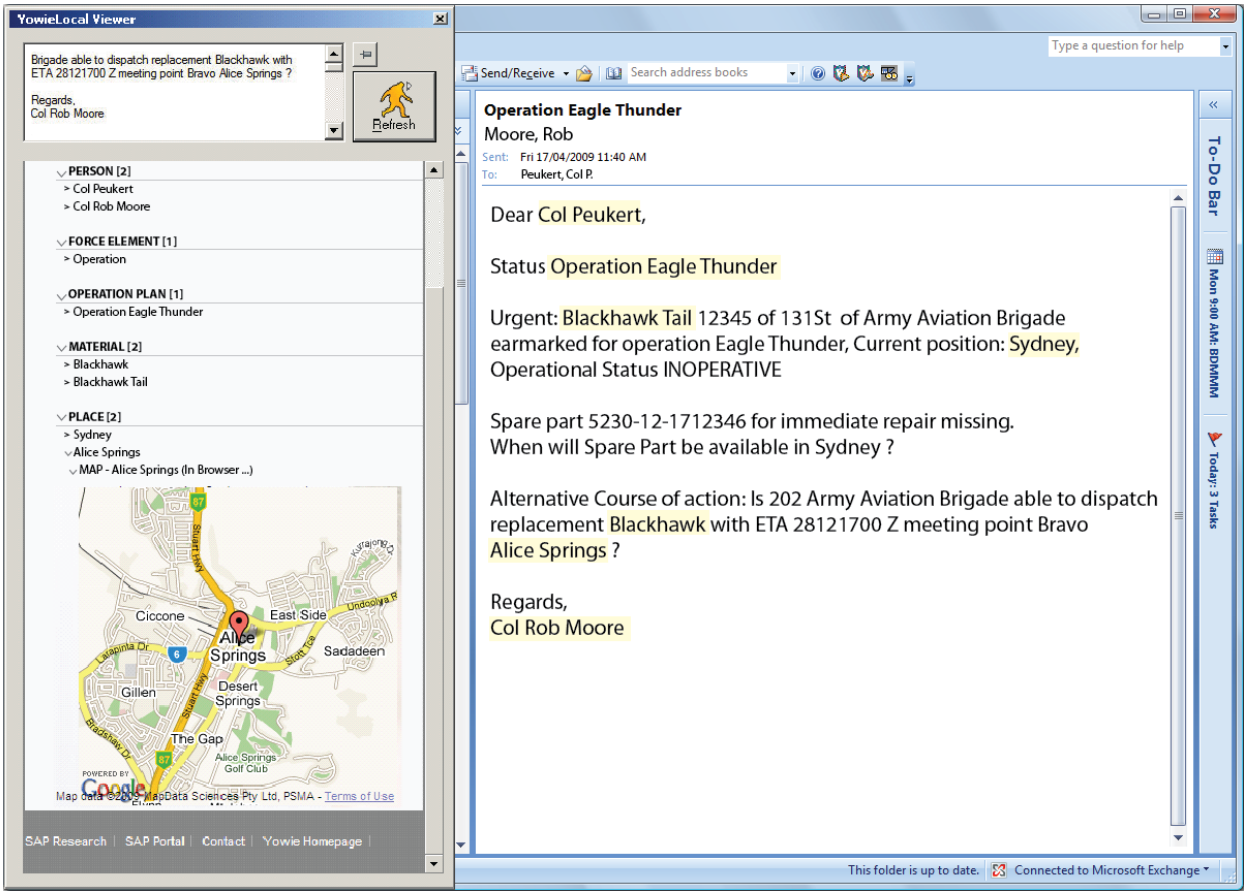
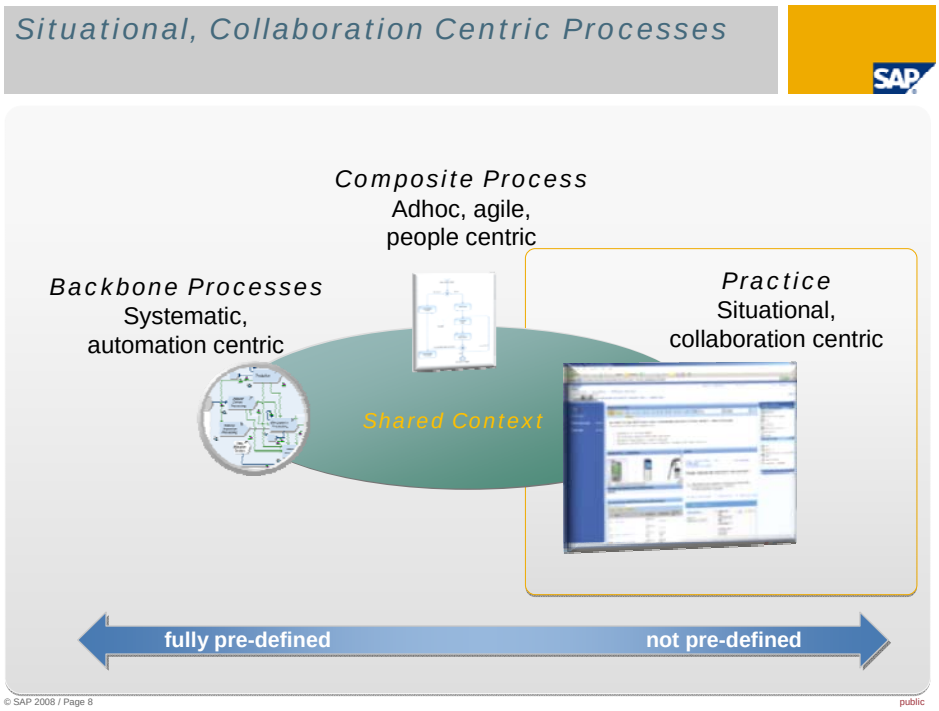


Figure 2 Yowie in a defense scenario

Through this prototype we are bridging from the unstructured email into the structured context of backbone process. In a nutshell this is a way to Transition from Isolation to Integration. These defense backbone processes can embrace the full spectrum of defense information systems.

The access fully relies of a fully implemented secure & trusted collaboration.

Ideally in the future Yowie would not only take into account Identifying Entities and Related Objects but also **make recommendations based on the User.** It should **recommend** to identified Entities related Objects based on your Role, for example peer-behavior. We are just starting investigating into that human element and hoping to see much more advanced scenarios soon.



From all of the attributes of the proposed approach, we can then map them to the developed C2 maturity model [4]. Yowie creates deep shared understanding through the automation of processes by identifying entities and related objects pertinent. Ideally in the future Yowie would not only take into account Identifying Entities and Related Objects but also make dynamically adaptive recommendations based on the User Role and peer behavior leading to shared understanding. We purport that this approach leads to "position 5" **– shared understanding and dynamically adaptive – in the C2 maturity model**.

5. Conclusions

The prototype is currently tested internally at SAP Research, and further work will include extending capabilities of the system to recognize more types of entities relevant for researchers. In case where entity recognition is ambiguous, we are currently working on methods ranking suggested entities.
A claim that information quality is being improved is currently based on an expert analysis. This is clearly not satisfactory, and further user studies need to be performed to ensure that the information quality, and therefore credibility is improved.

6. References

1. Addressing information overload in corporate email: The economics of user attention. Technical report, The Radicati Group, Inc, 2007.
2. D. Maynard, V. Tablan, C. Ursu, H. Cunningham, and Y. Wilks. Named Entity Recognition from Diverse Text Types. Recent Advances in Natural Language Processing, 2001.

3. U. Nahm and R. Mooney. Text Mining with Information Extraction. In Proceedings of the AAAI 2002 Spring Symposium on Mining Answers from Texts and Knowledge Bases, 2002.
4. David S. Alberts and Richard E. Hayes: Planning Complex Endeavors. CCRP, pp. 159-179, (2007)