



BlueSpace Mail Relevance

Contents

1.1 Index

1.1	INDEX.....	2
1.2	LIST OF FIGURES.....	3
1.3	LIST OF TABLES.....	3
2	EXECUTIVE SUMMARY.....	5
3	INTRODUCTION.....	6
3.1	EMAIL.....	6
3.2	EMAIL SEARCH.....	6
3.3	BLUESPACE.....	6
3.4	BUSINESS BENEFITS.....	6
3.5	USER IMPACT.....	7
4	BLUESPACE TECHNOLOGY.....	8
4.1	BLUESPACE RELATIONAL ARCHITECTURE.....	8
4.2	TRANSMAIL AND SUBMESSAGE PATHS.....	8
4.3	BLUESPACE EXTERNAL TRANSMAIL.....	10
4.4	BLUESPACE GOVERNANCE.....	11
5	BLUESPACE TRANSMAIL RELEVANCE.....	13
5.1	INTRODUCTION.....	13
5.2	MAIL DATA.....	16
5.2.1	Mail data inputs.....	16
5.2.2	Mail data relevance.....	17
5.2.3	Mail data collaborative relevance.....	17
5.3	MAIL ACTIONS.....	18
5.3.1	Mail actions inputs.....	18
5.3.2	Mail actions relevance.....	18
5.3.3	Mail actions collaborative relevance.....	20
5.4	MAIL ACTIONS IMPORTANCE.....	20
5.5	CORRELATED ENTITIES.....	21
5.6	EXTENSIONS.....	22
5.7	USER INTERACTION.....	22
5.8	SUMMARY.....	23
6	EMAIL COMPARISON.....	24
6.1	INTRODUCTION.....	24
6.2	EMAIL UNIT.....	24
6.3	RELATIONAL RESTRUCTURING.....	24
6.4	EMAIL INFRASTRUCTURE INTEGRATION.....	25
6.5	SUMMARY.....	26
7	USER INTERFACE.....	27
7.1	INTRODUCTION.....	27
7.2	BLUESPACE TRANSMAIL VIEW.....	27
7.2.1	Mail view.....	27
7.2.2	Folder view.....	33
7.2.3	Address book.....	33
7.2.4	Preferences.....	34
7.3	BLUESPACE TRANSMAIL SEARCH.....	36
7.3.1	Search form.....	36
7.3.2	Advanced search.....	38

7.3.3	Custom phrases.....	43
7.3.4	Search settings.....	43
7.4	BLUESPACE TRANSMail SEARCH RESULTS.....	45
7.4.1	Search results.....	45
7.4.2	Folder view.....	45
7.4.3	Extended view.....	47
7.5	SUMMARY.....	50
8	BLOGS.....	51
8.1	INTRODUCTION.....	51
8.2	RELEVANCE IN BLOGS.....	51
9	ALGORITHMS.....	52
9.1	BRAIN DUMP.....	52
9.2	INITIAL MATHS.....	53
9.2.1	Introduction.....	53
9.2.2	Importance.....	54
9.2.3	Example.....	55
9.2.4	Time-dependent relevance.....	55
9.2.5	Notes.....	56
10	CONCLUSION.....	57
11	GLOSSARY.....	58

1.2 List of Figures

Figure 1 – Email versus TransMail.....	8
Figure 2 – BlueSpace TransMail scenario.....	9
Figure 3 – Branching TransMail paths.....	10
Figure 4 – TransMail to external contacts.....	11
Figure 5 – Complete compliance model.....	12
Figure 6 – Relevance overview.....	14
Figure 7 – BlueSpace email integration.....	25
Figure 8 – BlueSpace TransMail view.....	28
Figure 9 – Asking the user to rate content relevance.....	30
Figure 10 – BlueSpace TransMail user view.....	31
Figure 11 – Importance evolution.....	32
Figure 12 – BlueSpace TransMail folder view.....	33
Figure 13 – BlueSpace address book.....	34
Figure 14 – Tuning the algorithm.....	35
Figure 15 – Simple search interface.....	37
Figure 16 – Regular search interface.....	38
Figure 17 – Advanced search interface.....	39
Figure 18 – Search settings.....	44
Figure 19 – Folder view search results.....	46
Figure 20 – Extended view search results.....	47

1.3 List of Tables

Table 1 – TransMail view icons.....	29
Table 2 – Message details icons.....	32
Table 3 – Algorithm tuning icons.....	35
Table 4 – Advanced search icons.....	40

<i>Table 5 – Search settings icons</i>	44
<i>Table 6 – Folder view icons</i>	46
<i>Table 7 – Extended view icons</i>	48

2 Executive Summary

E-mail was invented in the early 1970s as part of an academic experiment.¹ According to META Group, 75% of corporate knowledge will reside in email by 2005.² Today, business runs on email – it gets used for the vast majority of interactions between people.

Email has catalysed an increase in the speed of business and the volume of information presented to employees. Whilst an average employee five years ago might have received 30 letters and memos each week, many employees today receive 200 emails each day. PWC in the US has 25,000 employees, and recently indicated to BlueSpace that it handles 5 million emails each day.

The challenge for employees is to navigate through this 'information torrent'. To discover what is actually relevant to them through a sea of noise. This problem is particularly acute for management employees.

BlueSpace employs a relational architecture to add structure to mail content. Each piece of content is audited and managed through its lifecycle in the mail system, and every significant user action is audited. This provides a powerful set of raw data that can be used in relevance algorithms.

Most email systems today have basic search facilities in which each email is given equal weighting. BlueSpace allows searching by keyword, process code and other such criteria. But in addition, BlueSpace uses a multi-dimensional relevance system – BlueSpace Mail Relevance – which rates the importance of each piece of content based on the following criteria:

1. **User actions** – what did the user do with the mail? If they moved it to a folder or forwarded it then it is likely to be of significant importance.
2. **User collaborative actions** – what did other users do with the mail? Again, their actions indicate their views on the content's importance.
3. **Administrative settings** – a 'Relevance Administrator' can add weightings reflecting the enterprise's view of what is important.

BlueSpace Mail Relevance does not only rate content – it also rates the importance of each user and other entity (such as Jobcodes and contexts) in the system. A user can interact with the relevance and importance results by changing a 'slider' (graphical element) on a mail by mail or per sender basis, to help the system learn.

BlueSpace Mail Relevance provides much needed innovation in the user experience of electronic mail. A user can display a folder, but view only content of say 80% relevance and above, then handle the most important mail content first. The search facilities are far in advance of any comparable system. A user can change the relevance algorithm in real-time to improve search results – e.g. "show me content that was read by more people". A user can also view what happened to mail content they have written during its lifecycle, including recipients, read times and the path the content took.

Note that both users and administrators do not need to interact with the BlueSpace search algorithm (e.g. changing relevance/importance) for the system to work. It functions automatically based on what users do – though users can provide feedback to help the algorithm learn.

Email has become both the largest unstructured information store and the most comprehensive repository of knowledge in modern corporations. Enterprises and their employees need more effective mechanisms for mining the information in their mail systems. BlueSpace Mail Relevance provides such a mechanism. The granularity of the BlueSpace relational mail architecture allows access to the raw data required to deploy a highly relevant search system. BlueSpace allows people to find the content they're looking for in mail.

3 Introduction

3.1 Email

Email was invented in the early 1970s as part of an academic experiment.³ While its fundamental protocols (such as Simple Mail Transfer Protocol – SMTP) have changed little since their inception in 1982,⁴ real-world usage of email has changed dramatically. Email is now a mission-critical business tool, and one of the most resource-hungry services provided by IT. Analyst firm Gartner has indicated that the volume of email is growing by 50% every year,⁵ causing companies' bandwidth requirements to double every six months.⁶

SMTP treats each email as a string of text similar to a postcard or telegram that must be passed between computers until it reaches its desired recipients. When a user replies to or forwards on an email, they photocopy all the content in that email and put it into a new email text string with their own comments. This has made email systems into a 'content dumpster', with no structure or ability to audit content, or to mine the information inside the email system.

According to META Group, 75% of corporate knowledge will reside in email by 2005.⁷ But there is no way to mine this data effectively. This is particularly important for email that constitutes formal business records. Most corporations allow users to view only the mail content they have received directly, whilst traditional content management systems have access rights layered over the entire content store. This again reflects the history of email as a person-to-person communications system, rather than a system designed for knowledge management.

3.2 Email Search

Email search is a disparate market, with different products offering varying value propositions. Some companies offer a query-like store as a way to remove content from an active mail server to help such servers scale. Others have attempted to branch into email search from content management systems, but have faced challenges due to the lack of structure in the mail system and the lack of defined access rights. If a corporation would just allow everyone access to all emails, it would be carnage.

3.3 BlueSpace

BlueSpace uses its unique relational technology to provide greater granularity into content in a mail system. By coupling this to its governance model, BlueSpace can determine which pieces of a content a user should be allowed to search. The innovation described in this white paper involves adding advanced analytics into a search system in BlueSpace, causing a dramatic improvement in relevance for the end user attempting to find content.

This does not only apply to search, however, but also helps the user to get better at 'email browsing'. Many users in corporations are now at the sensible limits of information exposure through email, receiving up to 500 emails each day. BlueSpace email relevance is designed to help deal with this problem.

3.4 Business Benefits

The concepts and technology described in this white paper will benefit all businesses and organizations to some level, but there is one category for which it may become transformational – research. Organizations that have a

focus on research, particularly collaborative research, have many systems that attempt to leverage knowledge at the enterprise level by encouraging information sharing between individuals. But the user community tends to gravitate towards email, and rarely complies with rigorous document management policies due to the sheer volume of email content

Specific industry examples include:

- Pharmaceutical firms, in which researchers have focused scientific specialities, but need to collaborate in cross-specialist teams.
- Financial research functions, in which analysts produce reports for customers who subscribe to their distribution lists.
- Intelligence agencies, that generate large volumes of intelligence data, and then need to distribute this data to 'customers' while also rating its likelihood of truth or some other criteria based on real-time user corroboration.
- Investigatory agencies, such as a tax inspection service, who investigate individuals and organizations for malpractice, and desire better ways to find the 'needles in the haystack' through correlative inspection.

More general benefits and uses for the concepts and technology described in this white paper include:

- Recognizing Spam and decreasing the volume of users who waste time dealing with it – this applies to both external 'advertising' mail and internal 'noise' mail.
- Discouraging dysfunctional mail behaviour by end users, in which people unnecessarily copy many recipients on mails they write.
- Facilitating shared inbox efficiency (e.g. between managers and their secretaries).
- Allowing users to deal with the most important content first, e.g. when they come back from holiday and have 2000 mails in their inbox.
- For mobile users, only pushing the content by default that is judged to be the most relevant to them.
- Improving overall efficiency by helping users to better manage their mail 'information torrent' – a process that a typical user carries out for between two and five hours each day.

3.5 User Impact

Much of the innovation in previously published BlueSpace white papers has focused on either enterprise governance and control, or infrastructure efficiency or automation. An obvious question to ask is, "Will people use, adopt and 'fall in love with' the BlueSpace application?" This white paper outlines a powerful addition to the user experience – greatly improved and collaborative search capabilities that help the user navigate their mail. Email may have become the enterprises' core knowledge management repository. But it is ironic that the current, mainstream search facilities for end users in email applications such as Microsoft Outlook are so antiquated and ineffective.

4 BlueSpace Technology

4.1 BlueSpace Relational Architecture

BlueSpace has developed a messaging platform that is very different from existing messaging and mail systems. As introduced in Section 3.1, email photocopies content with each generation of reply and forward, as shown in Figure 1.

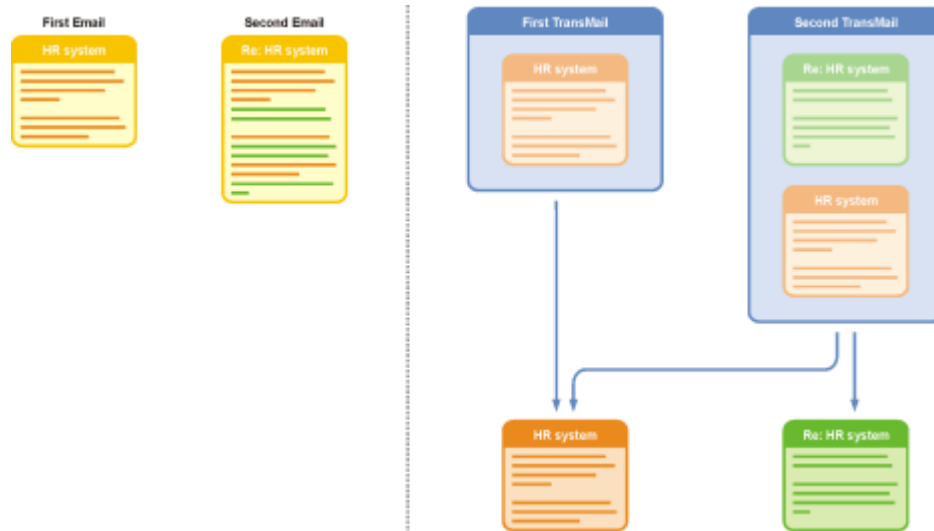


Figure 1 – Email versus TransMail

In the second email, the red message is duplicated without any reference to its origin – there is no link between the first email and the second email. In the BlueSpace TransMail system, the first TransMail points at the red message, and the second TransMail points at both the red and green messages.⁸ Both red and green messages are termed ‘submessages’ in the BlueSpace environment. There is a glossary defining all terminology in this white paper in Section 11.

This relational pointing has two powerful outcomes:

1. **Control:** It is far easier to manage and audit the content, enforce access rights and guarantee retention policies are rigorously applied, as there is just a single copy of the message to manage.
2. **Cost saving:** The TransMail system uses far less infrastructure than regular email systems, as it does not photocopy the same piece of content in to many separate places on the infrastructure.

4.2 TransMail And Submessage Paths

An important concept in the world of BlueSpace is that of paths – the way information is moved around a BlueSpace system. Due to the relational architecture of the BlueSpace TransMail application, BlueSpace is able to track what happens to content during its lifecycle, as well as enforcing rules over what is allowed to happen to it.

This section is included to assist the reader in understanding the relational model of BlueSpace, and its importance in search relevance. The user is not required or expected to go to this level of detail when using the application – it’s irrelevant to them, they just send mail.

Figure 2 illustrates an example of a series of BlueSpace TransMail communications in which the user Justin Marston is involved.

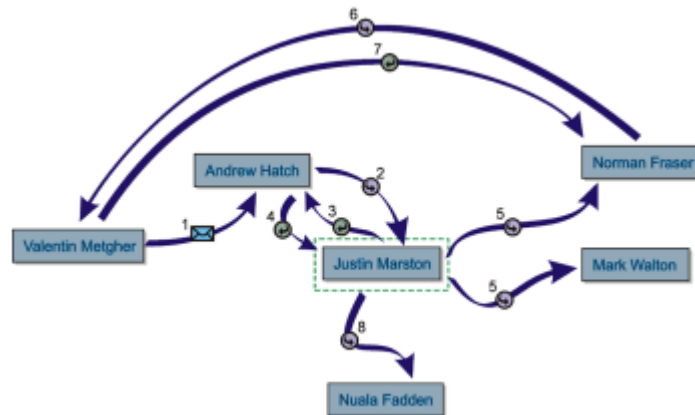


Figure 2 – BlueSpace TransMail scenario

The steps 1 to 8 in Figure 2 are as follows:

1. Valentin Metgher sends a TransMail to Andrew Hatch that contains submessage A.
2. Andrew Hatch forwards the TransMail to Justin Marston, creating a new TransMail that contains B|A (a new submessage B, with history portion submessage A).
3. Justin Marston replies to Andrew Hatch, creating a TransMail that contains C|BA.
4. Andrew Hatch replies to Justin Marston, creating a TransMail that contains D|CBA.
5. Justin Marston forwards the TransMail generated in step 4 to Norman Fraser and Mark Walton, creating a TransMail that contains E|DCBA.
6. Norman Fraser seeks to clarify the situation with Valentin Metgher, sending a TransMail containing F|EDCBA.
7. Valentin Metgher replies to Norman Fraser, creating a TransMail that contains G|FEDCBA.
8. Some time later, Justin Marston forwards the TransMail he received from Andrew Hatch to Nuala Fadden, creating a TransMail containing H|DCBA.

There are three types of submessage with respect to any given TransMail or other submessage:

1. **Current submessage:** The submessage at the top of the chain in any given TransMail is termed the 'current submessage', which will have been written by the user who created the TransMail. In Figure 1, the red submessage is the current submessage in the first TransMail, and the green submessage is the current submessage in the second TransMail. In step 5 of Figure 2, submessage E is the current submessage.
2. **History submessage:** The previous submessages in the chain of any given TransMail (e.g. the red submessage in the second TransMail in Figure 1) are termed 'history submessages'. History submessages are content that was created before the current submessage and its respective TransMail. In step 5 of Figure 2, submessages D, C, B and A are the history submessages.
3. **Future submessage:** For any given submessage, submessages that appear in later TransMails that reference the original submessage are termed 'future submessages'. Thus in Figure 1, the green submessage is a future submessage with respect to the red

submessage. In step 5 of Figure 2, submessages F and G are the future submessages, but not submessage H, as the TransMail generated in step 8 does not include submessage E. H is a future submessage in step 4 of Figure 2.

Note that a submessage can only have one history path, as it can be the 'current submessage' in only one TransMail. A submessage can, however, have many future paths, as it may be included in many branching forward and reply structures. This is illustrated in Figure 3.

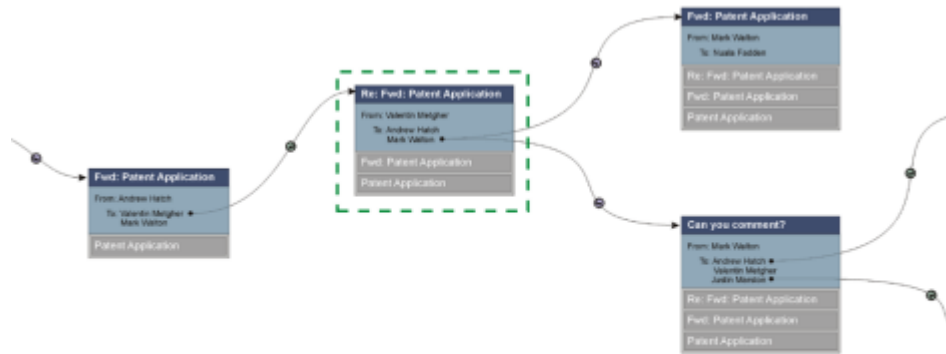


Figure 3 – Branching TransMail paths

All definitions of history and future are thus relative to a given point in time. In Figure 3, the ringed TransMail from Valentin Metgher to Andrew Hatch and Mark Walton is being considered as current. History submessages were written as part of the TransMails to the left, Future submessages were written as part of the TransMails to the right.

Note that a user may have already read content that constitutes both history and future submessages as part of other TransMails when reading a given TransMail.

4.3 BlueSpace External TransMail

BlueSpace can allow TransMail users to send content to external contacts, as shown in Figure 4.

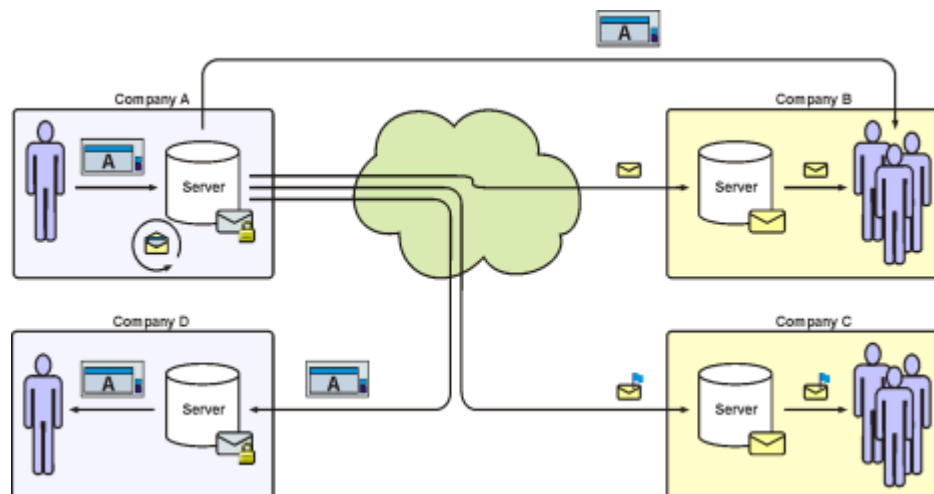


Figure 4 – TransMail to external contacts

Figure 4 illustrates three different methods:

1. A user in Company A sends a TransMail to users in Company B. The BlueSpace server in Company A sends an email notification (with a URL link to the TransMail) to the recipients in Company B. The recipients in Company B then click the URL to read the TransMail through a web browser.
2. A user in Company A sends a TransMail to users in Company C. The BlueSpace server in Company A wraps the TransMail up in an SMTP envelope, and sends it like a regular email to the email server in Company C. The recipients in Company C view it as normal. (Note that some control is lost by Company A, as it is unaware of the actions of users in Company C.)
3. A user in Company A sends a TransMail to a user in Company D. The BlueSpace server in Company A realises that the mail server in Company D is a BlueSpace server using BlueSpace Discovery Service.⁹ The BlueSpace server in Company A sends the TransMail to that in Company D using a protocol based on web services, and the recipient in Company D then reads the TransMail.

The second scenario, in which BlueSpace wraps a TransMail in an SMTP envelope, is the most complex to manage at the application level. Sending out the email is easy enough, but interpreting any replies can be a challenge, especially in corporations with BlueSpace systems deployed alongside regular email systems. This has been discussed in detail in the BlueSpace Email Structuring white paper.¹⁰

4.4 *BlueSpace Governance*

BlueSpace manages content during its entire lifecycle according to rules laid down in the BlueSpace Integrated Governance Management¹¹ model. This determines how long content is kept for, who can access it, where the data should reside, and the other allowable actions to be carried out by users.

A UML representation of the complete BlueSpace compliance model is presented in Figure 5.

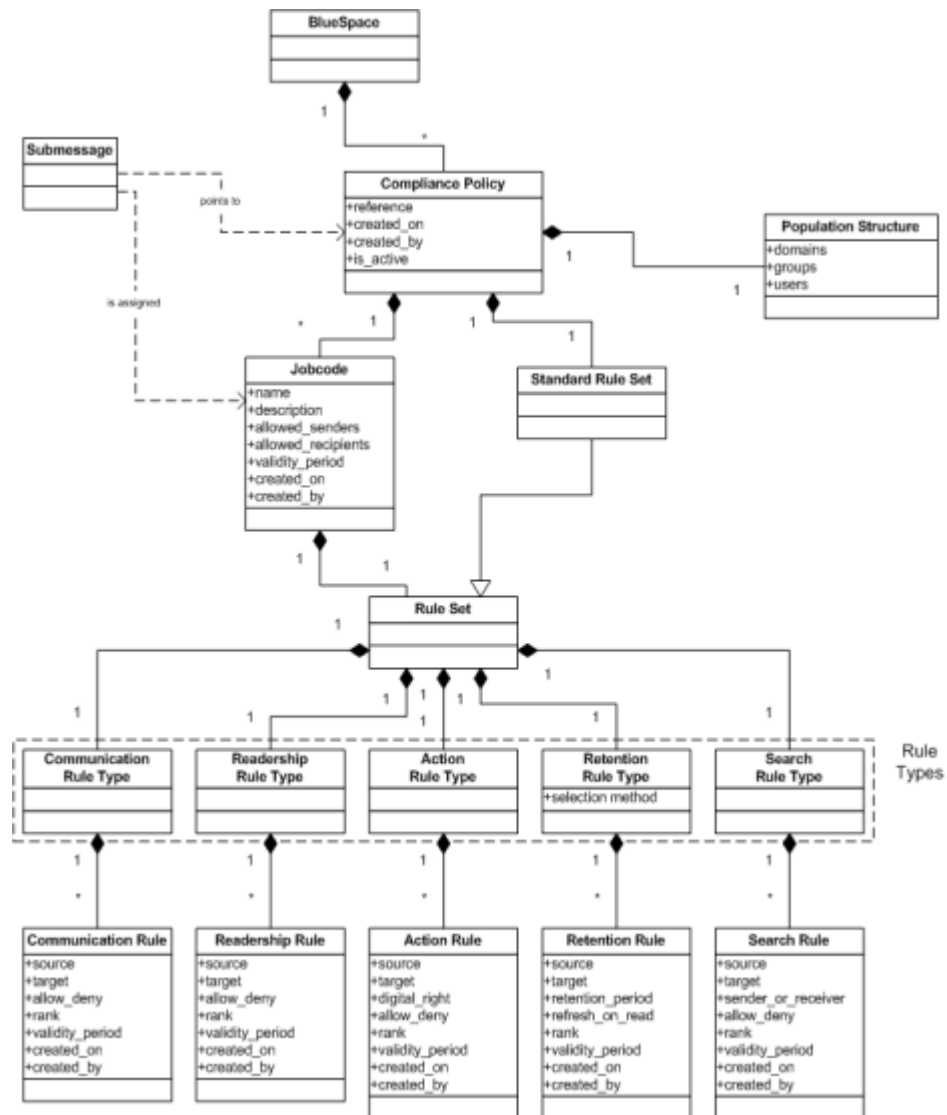


Figure 5 – Complete compliance model

The key aspect for this white paper is the ability of BlueSpace to layer search rules over a content system, determining who has the right to search different pieces of content in the system. For example, in an investment bank, an analyst might have the right to search all content related to analyst processes written by other analysts. However, the same analyst will not have the ability to search any content that does not pertain to his/her role in the bank.

The BlueSpace search rules technology is fully described in the BlueSpace Integrated Governance Management¹² model.

5 BlueSpace TransMail Relevance

5.1 Introduction

As described in Section 4.1, BlueSpace separates each TransMail into its constituent message objects, written by different authors. This provides far greater granularity than any other messaging system. The granularity and ease of access to the information helps to support the powerful analytics that BlueSpace can apply to a TransMail system. Behavioural information can be easily gleaned from the mail environment and used to help improve content relevance.

BlueSpace TransMail Relevance helps users as they perform two different actions:

1. **Browsing:** In this mode, the user moves through a folder of email content, such as looking through their new TransMails in their inbox.
2. **Searching:** In this mode, the user is actively looking for a specific piece of information, or for information of a given type.

Note that relevance scores for a given piece of content are absolute when browsing (as there are no specific criteria), but vary when a user is searching. For example, a piece of content is more relevant to a search if two of the words in the subject are keyword search terms than if there is only one. The user interface distinguishes absolute (browsing) relevance indicators from relative (searching) relevance indicators via colour, as outlined in Section 7.

There are two types of information that can be recovered from the mail system:

1. **Mail data:** Data in the mail itself, including keywords, creation dates, recipient information, Jobcodes and other meta-type data.
2. **Mail actions:** Information regarding users' actions on mail content, including the number of times it has been read, the number of times it has been forwarded, etc.

There is a glossary defining all terminology in this white paper in Section 11. In addition, end users can themselves rate various aspects and entities in the system to improve relevance, and can view the entire lifecycle of mail content they have written.

Users interact with relevance in three different ways:

1. Reading new mail – when a user receives a new TransMail, relevance can indicate how important that content is likely to be.
2. Reading existing mail – when a user is viewing their folder listing or a specific TransMail in the mail view, relevance can indicate how important each TransMail and submessage is likely to be.
3. Searching mail – when a user is searching through TransMail to find something, relevance can indicate how likely a given piece of content is to match the user's criteria, and can also act as a basis for ranking the search results.

An 'importance' factor is determined for each entity (users, groups, Jobcodes etc) in the system, acting to weight various inputs to the relevance algorithms. Importance is based on historical learning in the system, based on previous actions by users.

A high-level illustration of the various elements and relationships discussed in this white paper is presented in Figure 6.

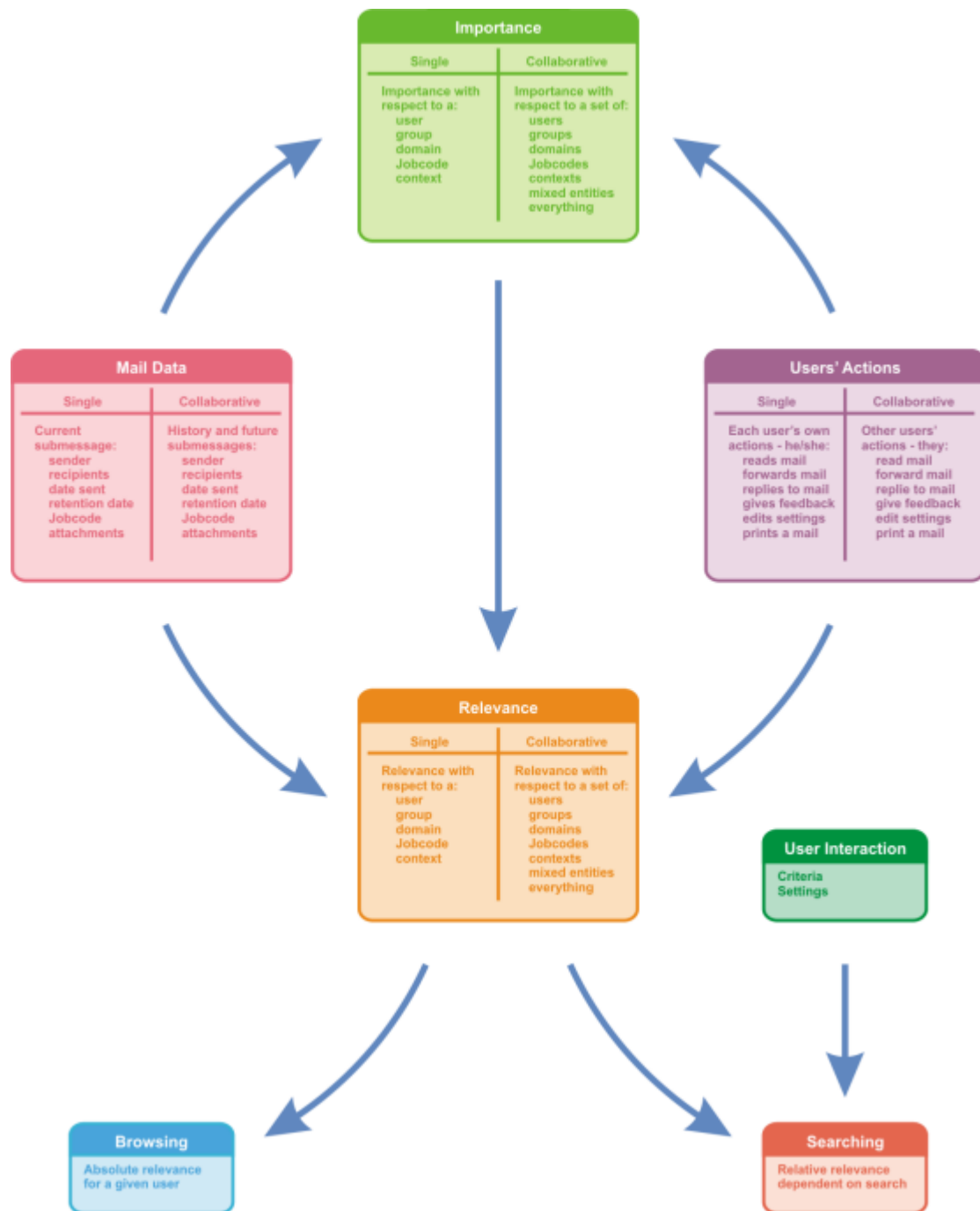


Figure 6 – Relevance overview

Figure 6 is clearly quite complex, but sums up the overall information flow in the BlueSpace relevance engine. Each aspect in the diagram is described below:

- **Mail Data:** Mail data is properties of each mail, including both the current submessage (new content) and history submessages (old content). Mail data acts as an input to importance, in that it determines the entities listed in the mail environment and where they appear. Mail data acts as an input to relevance, because the entities involved in the mail help to determine its initial relevance value.
- **Users' Actions:** Users' actions are the ways users interact with the BlueSpace system, e.g. reading mails, replying to mails and changing

their settings. Users' actions encompass a given user's own actions (a single user) and other users' actions (collaborative users in the system). These actions help to determine relevance, as relevance is calculated based on the way users interact with content. Users' actions also help to determine importance, as the importance of entities reflects the significance with which users treat them.

- **Importance:** The importance of a given entity can be stated with respect to a single entity (e.g. "how important is John to Douglas?") or a set of entities (e.g. "how important is John to the Calucite drug development team?"). Importance draws input from mail data, which determines the 'fields of engagement' for entities, and from users' actions, which reflect how significant a given entity is to a given user or set of users – 'the way the battle plays out'. Importance also acts as an input to relevance, as it helps to determine the relative weightings given to different actors in the system. Note that importance values for different entities are shown in a user's graphical interface (as described in Section 7).
- **Relevance:** The relevance of a given piece of content can be stated with respect to a single entity (e.g. "how relevant is it to John?") or with respect to a set of entities (e.g. "how relevant is it to the Calucite drug development team?"). Relevance has as inputs the mail data (properties of the content in question), users' actions (what users have done with the content), and importance (the significance users attribute to the entities involved in the content). Relevance outputs to two user activities – browsing (helping users navigate their mail folders) and searching (helping to determine the relevance of content to a specific search). Relevance values are displayed for content, helping users navigate through their mail folders.
- **User Interaction:** When users are searching for something, they have some idea of what they are looking for, and can guide the system to that content by stating criteria and changing their search settings.
- **Browsing:** This is the activity users spend most of their time doing in mail systems – navigating through mail folders and reading content. Relevance acts as an input here, helping to guide the user to the content they might wish to do deal with first.
- **Searching:** This is the activity that often most frustrates end users – not being able to find the mails they are looking for amongst the thousands of mails they have sent and received. The relevance of content to a given search varies depending on the search criteria settings (and user's settings).

There can be no doubt that this data schema is complex, as reflected by the computing and mathematics in Section 9. It may appear to be circular – and it is – because the relevance system in BlueSpace employs a feedback loop. It does have access to relatively independent data points, however, both by indirect 'watching' (users' actions and mail data) and by direct 'questioning' (asking users whether the results are accurate). These are only relatively independent, however, as users' opinions of the significance of both entities and content will be influenced by the values presented to them by the relevance system, and implicitly their trust in this system based on their views of its accuracy.

The concepts described above may seem rather abstract, so the following sections of this white paper outline 'real-world' examples that help to illustrate the way relevance and importance work in practice.

5.2 Mail Data

This section is divided into three:

1. Mail data inputs – both conventional and BlueSpace-specific criteria that can be used to search content.
2. Mail data relevance – scenarios illustrating how this data can affect search relevance.
3. Mail data collaborative relevance – scenarios involving other pieces of content in determining the relevance of a given piece of content.

5.2.1 Mail data inputs

Many of the criteria for searching mail data in BlueSpace are the same as those in traditional mail search environments:

- Keywords in subject
- Keywords in message body
- Keywords in other fields
- Keywords in attachments (both names and content)
- Any signature assigned to the message
- The date the content was created
- The recipients of the content
- Whether it was sent To:/Cc:/Bcc:
- The sender of the content
- The body size
- The attachment size
- The attachment type
- Flags for follow up – the sender indicates the TransMail requires action by the recipient.

Note that a user may build a search query looking for different submessages that make up a TransMail, and may specify different creation dates, senders, recipients and other details for each submessage they are looking for that make up a TransMail. The user interface supporting these highly specific queries is detailed in Section 7.3.2.

Other criteria are unique to BlueSpace, and leverage the relational architecture of the BlueSpace TransMail system coupled to the strong emphasis BlueSpace places on business process integration:

- Jobcodes – a mechanism for linking an entire submessage (as defined in Section 4.1) to a specific business process.
- Context – a mechanism for linking any text in a submessage to a specific process, either on a user or enterprise-wide basis, as shall be discussed in the BlueSpace Contexts white paper.
- Groups – sets of users across the system.
- Domains – an X.400-style naming directory across the system.
- Priority – an urgency tag linked to notification processes.
- Sensitivity – a confidentiality tag linked to readership and search rights.

- Validity date – the period of time for which the content is viewable by the user.
- Retention date – the period of time for which the content is retained by the corporation.
- Other Meta-style information stored relationally with respect to the mail.
- Other business processes that are triggered by the mail event.

5.2.2 Mail data relevance

Examples of ways in which these parameters might affect relevance to a given user as part of a search include:

- If the keyword they are looking for is in the body, then it's likely to be relevant to them.
- If a user is looking for a Microsoft Word file, then TransMails with word attachments are relevant to the search.
- If a user is looking for a piece of content they received some time ago from the HR department, which was pretty important, then all of those criteria (date period, priority and the HR group/domain) will be relevant to the search.
- If a user is looking for all content relating to a particular M&A transaction, and that content has been tagged with Jobcodes, then content with the appropriate Jobcode number will be relevant to the search.
- If a user received a TransMail as one of fifty other Cc: recipients, it is less likely to be of relevance to them than if they received the TransMail as the sole recipient.
- If a user is looking for all TransMails that will trigger a financial purchase over the next 30 days in a BlueSpace system deployed for ERP, then the user can search for TransMails that will cause a purchase over the defined date range.
- If a mail is flagged for follow up by the sender, it is more likely to be of significant relevance to the user.

5.2.3 Mail data collaborative relevance

Other scenarios leverage the relational structure of BlueSpace to determine the 'communication context' of the TransMail. These scenarios rely on both history submessages (older submessages in the chain of the TransMail), and future submessages (submessages written later as part of other TransMails that contain the submessage), as discussed in Section 4.2.

- If a TransMail is received by a user as a reply to a TransMail they themselves sent, then a submessage in the history will map to the 'current submessage' (the one at the top of the chain) in a TransMail in their sent folder. In this case, the incoming TransMail is likely to be of significant relevance.
- If a particular submessage appears many times in a user's inbox, then a new TransMail containing this submessage is likely to be of relevance to the user.
- If a new TransMail received by a user contains one or more history submessages written by the user, then it is likely this new TransMail will be of relevance to the user.

- If historical submessages in a new TransMail were found to have high relevance ratings, then it is likely that the new submessage will have significant relevance.
- If a submessage, A, does not contain a given keyword, but many submessages that were sent in replying to and forwarding submessage A do contain the keyword, then submessage A is likely to still be relevant in a search for the given keyword.

Note that these cases do imply user actions, and thus overlap with those given in Section 5.3.

5.3 Mail Actions

This section has been divided into the following subsections:

1. Mail actions inputs – the user actions recorded by the system.
2. Mail actions relevance – scenarios illustrating relevance based on the user's own actions.
3. Mail actions collaborative relevance – scenarios involving other users' actions to determine content relevance.

5.3.1 Mail actions inputs

BlueSpace audits all significant user actions inside a BlueSpace system. These actions are important as they reflect the users' opinions of the content, and the value they place on it. Example of such actions are:

- A user composes a TransMail.
- A user reads a TransMail.
- A user forwards a TransMail.
- A user replies to a TransMail.
- A user deletes a TransMail.
- A user files a TransMail in a custom folder.
- A user prints a TransMail.
- A user saves a local copy of a TransMail.
- A user copies a TransMail to their clipboard.
- A user links a TransMail (or part of a submessage) to a context.
- A user opens or saves an attachment.
- A user forwards an attachment.

5.3.2 Mail actions relevance

Mail actions relevance relies on the user's actions to help determine content's relevance for content in all the user's mail folders – both sent and received. Consider the following user action scenarios, in which a user's own actions can be used as an input to a relevance algorithm:

- If a user receives a TransMail from another user, and immediately forwards it to three other users with comments, then it is likely that the content had relevance to them.
- If a user receives a TransMail and replies to it, then the content of the original TransMail is likely to be of significant relevance.

- If a user receives a TransMail from another user, and immediately deletes it, it is unlikely that they perceived the content to be of significant relevance.
- If a user reads a TransMail and then moves it to a custom folder out of the inbox, then it is likely that the content had relevance to them.
- If a user has read a given TransMail many times, then it is likely to be of significant relevance to them.
- The longer the user has a TransMail on screen for, the more likely it is to be relevant to them.
- If a user copies text from a given TransMail to their clipboard in Windows, assuming they have the rights to do so, then this will likely be more relevant than a TransMail in which they did not do so.
- If a user links a piece of content to a context, it is more likely to be of relevance to them.
- If a user always deletes content from a specific user quickly, then it is unlikely that any content from this user is of significant relevance to them.
- If a user always reads content from another specific user before reading content from other users in any given new set of new mail received into their inbox, then it is likely that content from this specific user is of significant relevance.
- If a user typically only reads short TransMails rather than long ones, then the chances are that short ones are more relevant to that user than longer ones.
- If a user never reads messages sent to a particular group or domain (that they are a member of), then any content sent to that group or domain is likely to be of less relevance to the user.
- If a user never reads TransMails tagged with a particular Jobcode, then this content is likely to be of little relevance to them.
- If a user rarely reads TransMails designated as lowest priority (or reads them last when browsing their inbox), then lowest priority content is likely to be of little relevance to them.
- If a user sends a TransMail to a lot of users, it is likely to be more relevant than a TransMail they sent to a single user.
- If a user sends a TransMail, and then later forwards it to other users, it is likely to be of more relevance than a TransMail that was sent once but never resent.
- If a user reads a TransMail from their sent folder, then the content is likely to be of significant relevance.
- If a user saves an attachment to their local hard disk or network system, it is likely to be of significant relevance.
- If a user prints an attachment or TransMail, it is likely to be of significant relevance.
- If a user sends a TransMail to their boss, it is likely to be more important/relevant than a TransMail to a junior secretary.
- If a user sends a TransMail with extensive formatting, it is more likely to be important/relevant to them than if they write a few lines of plain text.

- If a user has not accessed the folder in which the content is stored for over a year, then it is likely that the content is not particularly relevant.

5.3.3 Mail actions collaborative relevance

Mail actions collaborative relevance relies on the actions of other users to determine the relevance of content in a given user's folders – both sent and received. In the following scenarios, other users' actions can be used to help determine the relevance of content to a given user:

- If a user sends a TransMail to someone, and they read it, it is more likely to be relevant (to both the receiver and sender) than if they do not.
- If a user sends a TransMail to a group of users, and most reply to it, forward it or copy it to custom folders, it is more likely to be relevant than if they all delete it immediately.
- If most other recipients of a TransMail delete it as soon as it arrives in their inbox, then the recipient in question will probably not find the content particularly relevant.
- If most other recipients copy a given TransMail to a custom folder (or reply or forward it) once they have read it, then the recipient in question will probably find the content to be of significant relevance.
- If a given user is always copying another user on their mail content (e.g. an employee copying their boss) then the content is less likely to be of relevance to that other individual.
- If a user sends a submessage to another user, but the recipient doesn't read it, then this may have an affect on the relevance of the content.
- If the vast majority of the users in a given department read a piece of content sent globally, then the content is likely to be relevant to another user in that department.
- If the majority of other users tend to read content linked to a particular Jobcode, then content attached to that Jobcode is more likely to be relevant to the user in question.
- If the majority of other users tend to read content linked to a particular context, then content attached to that context is more likely to be relevant to the user in question.
- If most users read content from a particular author (e.g. the CEO), then content from this author is typically of greater relevance than content from a different author that no-one ever reads.
- If most users reply to content from a particular author (e.g. the CEO), then content from this author is typically of greater relevance than content from a different author that no-one ever replies to.

5.4 Mail actions importance

In addition to considering content and users' actions on it, the application can also analyse the importance of particular actors and entities when determining their relative weighting in calculating relevance. Such entities include:

- Users
- Groups
- Domains

- Folders
- Jobcodes
- Contexts
- Signatures
- Other business process information
- Any back-end/automated actors (e.g. external computer systems)

In normal business operations, every employee has an importance rating for other employees. For example, it is likely that an employee would find other individuals who report to them and those to whom they themselves report to be of greater importance than individuals working in a different department.

Importance in BlueSpace is also a relative value, respective to each user (or other entity) in the system. For example, users, groups, domains and other entities all have different relative importance values describing a particular Jobcode. Ultimately it is possible to calculate an overall importance value for any given entity in the system, which describes that entity's importance to the entire enterprise (defined either as all users or all entities).

Importance for any given contact can in fact be a pair of different values. The first value reflects the user's interest in content they received from the contact, and the second value reflects the user's interest in content they sent to the contact. Whilst it might be expected that these values would be very similar, there are scenarios in which they would not be similar. If the head of HR is extremely 'chatty', he/she may have low importance when sending content to other users, but those other users may view the head of HR as being very important when they themselves are doing the communicating.

Varying weightings for relevance inputs with importance means that if a user such as a boss deletes a TransMail, then this will have a greater affect than if a junior secretary deletes it. Thus different users' actions have different affects on relevance, due to varying importance factors.

Importance has multiple feedback affects, as if user B receives a TransMail from user A, and then user B forwards it to user C who is of high importance (e.g. the CEO), then this affects the importance of all three users.

In an organisation with, say, 100,000 employees, it is likely that any given user will communicate with new contacts each week. Importance is a relative factor for each pair of users' (or entities') interactions, and is stored with respect to each user, as described above. However, when a new contact communicates with a user for the first time, the system can still make a 'best guess' about that new contact's importance based on collaborative importance. The value then changes as the user interacts with this new contact. This first 'best guess' value is important, however, as in a system with a large population, the usability would be hampered if every new contact started 'at par'.

In certain scenarios, an organisation may decide to tie importance to roles rather than specific users. This is particularly likely in highly regimented environments, such as the military, where the role is a greater determining factor than the actual human personality involved. The system is able to make special allowances when a user changes job, as their relevance to others will change quickly, as would be expected.

5.5 Correlated Entities

As discussed in Section 5.4, a user will often be interacting with a new entity for the first time in large system, and thus the initial values of importance are very important. Another approach to these initial values is to recognize correlated entities, described through tree structures.

For example, one HR employee is generally likely to have similar importance to another one, especially if they are of similar position in a tree structure – e.g. they both work in employee appraisal form processing. There are clearly exceptions to this trend, but in most organizations it is likely to have a positive affect on the initial states of importance.

Note that if required, branched structures – such as knowledge taxonomies for contexts, directories for users and Jobcode hierarchies – can be used as a correlative basis for importance on an ongoing basis. Thus for example, an importance change for one user would have a cascading ‘ripple effect’ for similar users.

5.6 Extensions

The relevance methods described in this white paper have focused on the users’ interactions with the mail system. It should be noted, however, that a corporation can decide to extend the action observations for relevance to the local file system. A user opening a file, editing a file, etc can then be made to impact the file’s score in the relevance engine of the BlueSpace system.

A second extension specific to files is that of correlating relevance values together for various versions of the same file, e.g. as part of a collaborative document creation process. This is facilitated by the BlueSpace Document Management system, which tracks various file versions through the creation process using deal room techniques.

5.7 User Interaction

In addition to the actions that users perform in the BlueSpace system, BlueSpace can take direct user input to help determine relevance. For example, a user can have a ‘importance slider’ for each user in the TransMail system. For example, a user might make their boss ‘very important’ but their junior, chatty sales clerk ‘rarely important’ – these weightings can then be added into the relevance algorithm.

Each piece of content (submessage) can be presented to the user along with a slider that indicates its current relevance score. The user can move this slider, and apply the change in relevance to that specific piece of content, or add that change back in to how relevance is calculated across the whole user’s mailbox.

To assist in providing tangible data points for the relevance algorithm, the application can also directly query the user on the relevance of content sampled from the system, as indicated in Figure 9. Note that this sampling mechanism can also be used as a way to corroborate data in an R&D or intelligence community. In the former case, researchers can be asked to rate the usefulness of a piece of content to the project they are currently working on. In this latter case, rather than asking the user to rate the relevance of a piece of content, the application asks the user to rate their estimation of whether the content is true. Sampling can be conducted in a random or structured manner.

In addition to the end user being able to affect relevance directly for their own inbox, certain ‘super-user’ roles have the ability to affect relevance across the entire system. In other words, a ‘Relevance Administrator’ can provide a boost to communications from the HR department. The reason for this is that relevance is not a purely user-centric phenomenon – a separate question relates to relevance for the entire enterprise. Most people may ignore interaction with the HR department, but that doesn’t mean that the executive board regard communications from the HR department to be irrelevant.

5.8 Summary

BlueSpace assists users in navigating the deluge of mail content they receive each day, by calculating relevance ratings for content. This mechanism also impacts searching. Users can base their searches on conventional inputs, but can also use new criteria if they wish. The application ranks search results according to a relevance algorithm, rather than affording each result equal weighting. In calculating the relevance value, the application also takes into account the importance of each user in the system.

6 Email Comparison

6.1 Introduction

BlueSpace TransMail has a relational mail architecture, in which each piece of content written by a user is managed as a single instance object during its life in the mail system, as described in Section 4.1. The granularity of this architecture allows the application to record all actions by users in the system, ensuring that the application has audited the entire lifecycle of the content.

While regular email does not have this relational structure, the relevance techniques described in this white paper can still be applied to email at different levels:

1. Email unit – each unit of email is considered as a whole, no matter what other message fragments (other emails) it contains.
2. Relational restructuring – each email is broken down into its component messages, and the relational structure of BlueSpace TransMail is mapped to the email store.

In addition to application to historical content, BlueSpace Mail Relevance can be applied to email based on ongoing user actions through interaction with the email infrastructure.

6.2 Email Unit

Pure email unit analysis is the most basic approach to relevance in email content. It allows relevance to be based on simple criteria such as:

- The number of recipients
- Whether the user in question received the email as a To:, Cc: or Bcc: recipient
- The importance flag of the email
- The time the email was sent

This approach provides the least accurate relevance information, as the inputs are very limited.

6.3 Relational Restructuring

Regular email does not have the relational architecture of BlueSpace – each time a user replies to an email they photocopy the history into a new text string, thus making it very difficult to audit the lifecycle of the mail content.

The BlueSpace application has the ability to analyse regular email content and convert it into a relational system, similar to BlueSpace. This was first introduced in Section 4.3, and has been discussed in detail in the BlueSpace Email Structuring white paper.¹³

This approach is less elegant than structuring mail at the front end of the process as in BlueSpace TransMail. When the BlueSpace application structures information once it has already been sent, the criteria for search are those such as:

- Keywords and other standard inputs
- User mail flow actions – compose, reply and forward
- Collaborative user mail flow actions – compose, reply and forward

6.4 Email Infrastructure Integration

When the BlueSpace application is integrated with the email environment at the application level (as opposed to being a purely 'downstream' process), then the following criteria are also available:

- User mail management actions – read, forward, reply, delete and file to custom folder, etc
- User attachment management actions – view, forward, save to disk, etc
- Mail content path – tying together multiple instances of the same email

By employing the relational architecture for regular email, BlueSpace allows powerful searching of regular email, with a similar relevance system to that of the BlueSpace TransMail search engine. BlueSpace also layers the access rights policy model over the search results, as described fully in the BlueSpace Integrated Governance Management¹⁴ model.

Integration with email systems can be achieved in two ways, which are not mutually exclusive:

- Client-side integration, in which a BlueSpace plug-in to the email client monitors end user's actions on the mail content.
- Server-side integration, in which a BlueSpace server connects to the email server and tags content with 'flags' so that it can track what happens to the content.

These two mechanisms are illustrated pictorially in Figure 7.

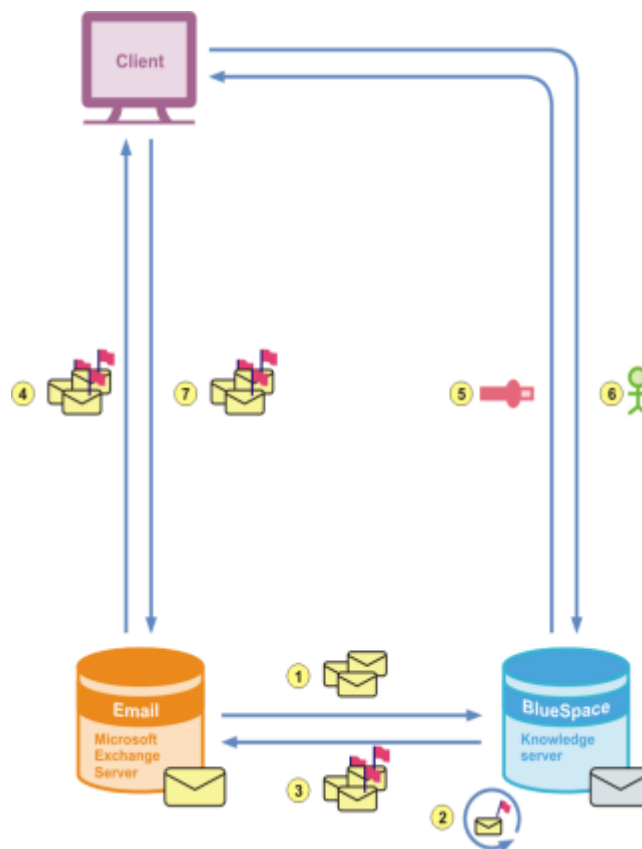


Figure 7 – BlueSpace email integration

The processes in Figure 7 are as follows:

1. New email received by the email server, e.g. Microsoft Exchange, is retrieved by the BlueSpace Knowledge Server.
2. The BlueSpace Knowledge Server adds meta-tags to each email, so that it can track that content through the lifecycle.
3. The tagged emails (represented by BlueSpace flags) are then replaced on the email server.
4. The email client, e.g. Microsoft Outlook, downloads the tagged emails.
5. The BlueSpace plug-in to the email client immediately retrieves the current relevance values for the content.
6. The user carries out actions on the content in the email client, and the BlueSpace plug-in records these actions and sends them back to the BlueSpace Knowledge Server.
7. One of the actions carried out by the end user is to send emails in replies and forwards – these include the meta-tags positioned by the BlueSpace Knowledge Server.

Note that steps 5 and 6 are a continuous feedback loop – each time the user carries out an action, the relevance value changes.

The BlueSpace plug-in to the mail client also interacts with search results, to include and reflect the relevance values in the search listings. These operations are similar to those described in Section 7.4 in the context of BlueSpace TransMail search results.

6.5 Summary

BlueSpace email search may be used as a standalone knowledge management system, for regular email systems with no TransMail penetration. It may also be used for searching legacy data once a mail system has been transferred to TransMail, or on external email content received through an SMTP gateway in a BlueSpace system with external email capabilities.

Note that to apply the BlueSpace mail search principles of relevance and user action indicators, the system does not need to restructure the content into the BlueSpace relational store. With client-side integration, BlueSpace is able to capture these actions at the application level rather than relying on inspecting mail data. BlueSpace then keeps a separate database recording user actions and links into the external content repository (mail server) through tagging mail content as users interact with it and conducting searches via a backend connector.

7 User Interface

7.1 Introduction

As first presented in Section 3.5, BlueSpace relevance and importance in mail search cause a significant improvement in the end user experience of a mail environment. BlueSpace mail search is semi-automated, in that users' actions determine the baseline behaviour of the system, but any given user can interact directly with the inputs into the relevance algorithm to better tune it to their own requirements.

This section is dedicated to describing the user interface for BlueSpace mail search, and more specifically, the BlueSpace TransMail search system when delivered through the default web interface.

7.2 BlueSpace TransMail View

There are various core pages of the BlueSpace TransMail web application that include relevance and importance aspects:

- Mail view – when viewing a single TransMail.
- Folder view – a folder listing showing many TransMails.
- Address book – a list of contacts, groups and domains.
- Preferences – a list of information

In addition, the administrative interface presents these statistics to end users.

7.2.1 Mail view

In the TransMail view, a user is looking at a single TransMail, and able to see its entire content. Various popup boxes and screen customisations are enforced on the mail view.

7.2.1.1 The reply view

An extension of this view is the reply view, as shown in Figure 8.

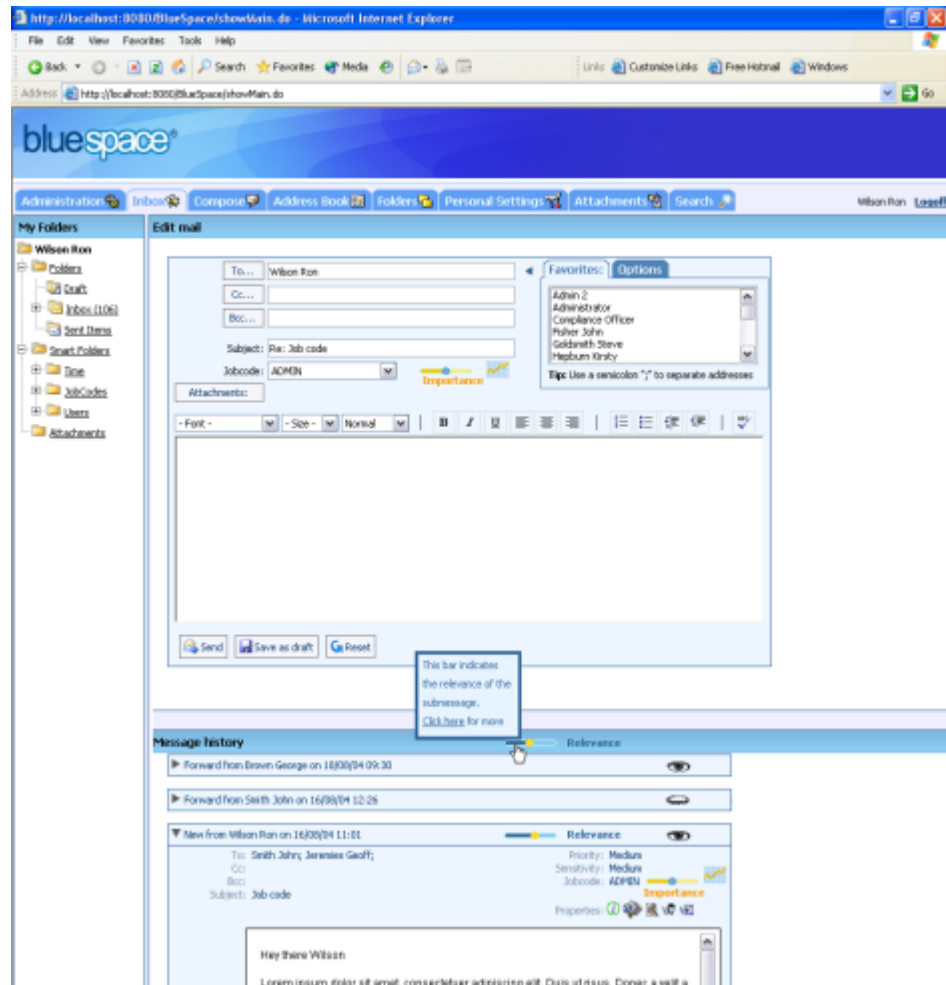


Figure 8 – BlueSpace TransMail view

There are various icons in Figure 8, and these are listed in Table 1.

Icon	ALT Text	Action
	(Popup explaining value)	Importance slider icon allows the user to change the importance value for the entity in question.
	Importance chart	Spawns the popup shown in Figure 11.
	(Popup explaining value)	Relevance slider icon allows the user to change the relevance value for the content in question.
	You have seen this submessage before	If submessage is shown collapsed, then expands it so it can be read. If submessage is shown expanded, then collapses it.
	You have not seen this submessage before	Expands the collapsed submessage.
	More information	Spawns the popup shown in Figure 10.

	Find all copies of this submessage	Shows a custom Smart Folder (dynamic) with all TransMails that contain the given submessage.
	Find submessages like this	Shows a search result page in the default preference of the user (see Section 7.4), with submessages deemed to be similar to the given submessage.
	Visualise users	Shows a popup with a visualisation of the user interaction that took place to make this TransMail, similar to Figure 2.
	Visualise mail	Shows a popup with a visualisation of the mail interaction that led to this TransMail, similar to Figure 3.

Table 1 – TransMail view icons

There are various relevance slides in Figure 8 – one for each submessage in the TransMail view, and one for the entire TransMail, taking into account the current submessage and all history submessages. Relevance sliders for history submessages are only shown when the submessage is expanded, as in the second history submessage in Figure 8.

When a user ‘hovers’ their mouse over a relevance bar, a popup appears, indicating why a particular relevance value is as it is – helping the user understand how the algorithm came to its conclusion. A similar popup appears for the importance sliders, discussed below.

The user can provide feedback to the algorithm, by changing the level of any given slider. When a user clicks and drags the level, similar to changing a graphic equalizer, the application shows a popup above the slider indicating the value to which the user has moved the slider. The respective values are:

- Irrelevant – 0% to 10%
- Slightly Relevant – 11% to 30%
- Relevant – 31% to 70%
- Highly Relevant – 71% to 90%
- Critical – 91% to 100%

The interface indicates to the user which submessages they have already seen before (e.g. in the case of a reply history), and which are new (e.g. in the case of a forward history). The interface shows this in Figure 8 via the state of the eye in each submessage. If the eye is open, the user has seen the submessage, if the eye is closed, the user has not seen the submessage. Thus in Figure 8, if the active user (Wilson Ron) opened the submessage from Smith John, the respective eye icon for that submessage would become open.

Figure 8 has further icons, which are described in Table 1. One icon allows the user to see a custom search folder (Smart Folder) showing every TransMail that contains the given submessage. Another searching icon allows the user to search for similar content to the given submessage, as determined by the search algorithm. Two more icons allow the user to view the lifecycle of any given submessage, spawning pop up windows that visualise its path through the system and all users that have carried out actions on it.

In addition to relevance feedback, the TransMail view can present information on importance. Jobcode importance and other custom created variables that are relatively sparse are given importance sliders as shown in Figure 8. A user can alter an importance slider, and either lock it at the new value or allow it to recalculate over time. A locked importance slider for user Smith John is shown in Figure 10.

7.2.1.2 Mail view relevance and importance sampling

In addition to being able to provide feedback to the algorithm by moving the relevance sliders, the user may also be requested to provide direct feedback on a sampled basis. In this scenario, the user is asked to directly rate the relevance of a given TransMail, thus presenting the algorithm with a direct relevance data point, as opposed to correlating relevance with the user's actions. This direct user questioning is shown in Figure 9.

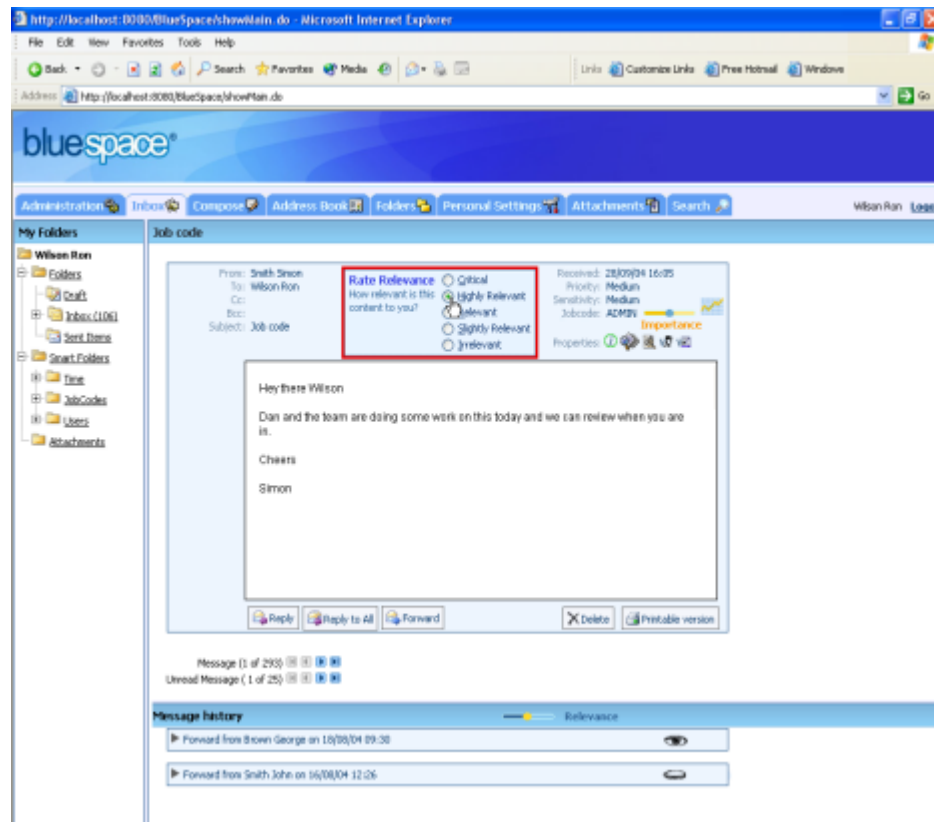


Figure 9 – Asking the user to rate content relevance

The user is not usually forced to set the relevance value before being able to leave the TransMail view, but the interface indicates that it is important for them to do so, and this would typically be reinforced by user training. It is important that a high response rate is achieved for these relevance rating questions, as the effectiveness of the algorithm correlates to the lack of bias in the survey sampling.

The number of times the user is asked to rank content can be determined based on their usage frequency (e.g. every 100 TransMail views) or absolute time (e.g. 5 times each week). A BlueSpace system administrator can determine this value, with the option to allow users to customise it through their preferences. An administrator can also make direct user interaction mandatory, for scenarios in which this is required (e.g. intelligence communities).

The direct feedback values created when users assign relevance to content through the survey function are most useful in the early stages of a given system. They help provide absolute relevance values to stabilise the weightings given to each type of user action, as discussed in Section 7.3.4. An administrator is therefore able to either select a time-based factor to decrease the frequency of surveying users. Alternatively, an administrator can base decreases in sampling frequency on statistical results derived from, for example, the stability of the global weighting factors.

Similar mechanisms are available for direct sampling of users' ratings of importance of other users. These also function by replacing sliders, in this case importance sliders, with a survey capability.

7.2.1.3 User mail actions

When it comes to users' actions and users' importance (as well as that of groups and domains), there's a lot of data. Rather than clutter the interface with all this information by default, the application allows the user to click the details icon (the green 'i' in Figure 8) that then displays a new box that presents this information. This is illustrated in Figure 10. It shows the sender and each recipient of the TransMail, their actions on the content (forwarding the TransMail, replying etc), and the importance of the each recipient.

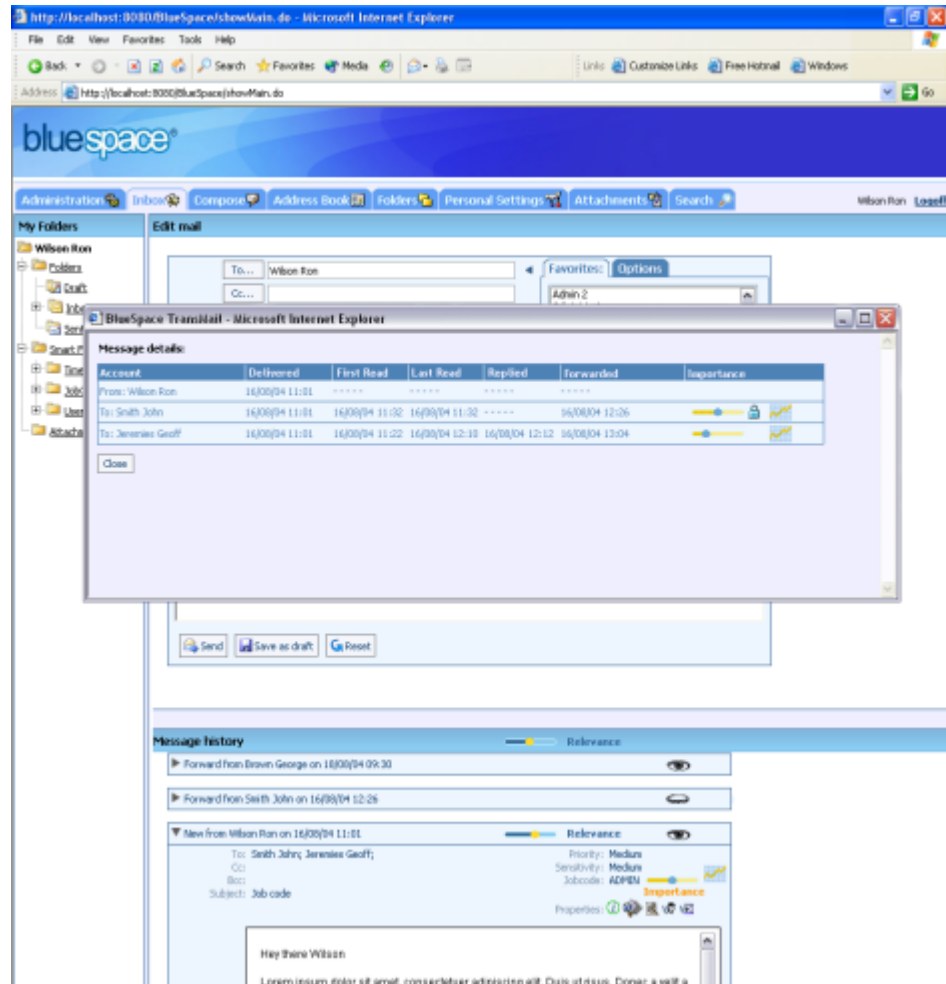


Figure 10 – BlueSpace TransMail user view

The icons in the message details popup shown in Figure 10 are described in Table 2.

Icon	ALT Text	Action
	(Popup explaining value)	Importance slider icon allows the user to change the importance value for the entity in question.
	Importance chart	Spawns the popup shown in Figure 11.
	Importance locked	Allows the user to unlock the importance value, so it can vary again.

Table 2 – Message details icons

Importance is again presented through sliders that the user can change – e.g. to tell the application that the user's boss is important. When a user makes a change, they can either lock the value with respect to the system (in order for the user to stay at that relative importance) or they can allow it to find its equilibrium value. The latter is like stressing a spring in a spring lattice – over time it will return to its lowest energy state, though this is likely to be different to the initial state as the system is continually changing over time. A locked slider is shown for Smith John in Figure 10.

7.2.1.4 Importance charts

Importance clearly changes over time – for example, when a user is interacting strongly with another user on a given project, then that other user will be of greater importance at that point than a year after the time when the project has finished. Similarly, if a user is working on a deal linked to a specific Jobcode, that Jobcode will be of far greater relevance during the deal than six months after the deal has finished. The chart icon in Figure 8 and those in Figure 10 can generate a chart showing the evolution of importance over time – for a definable period set by the user. This popup is presented in Figure 11.

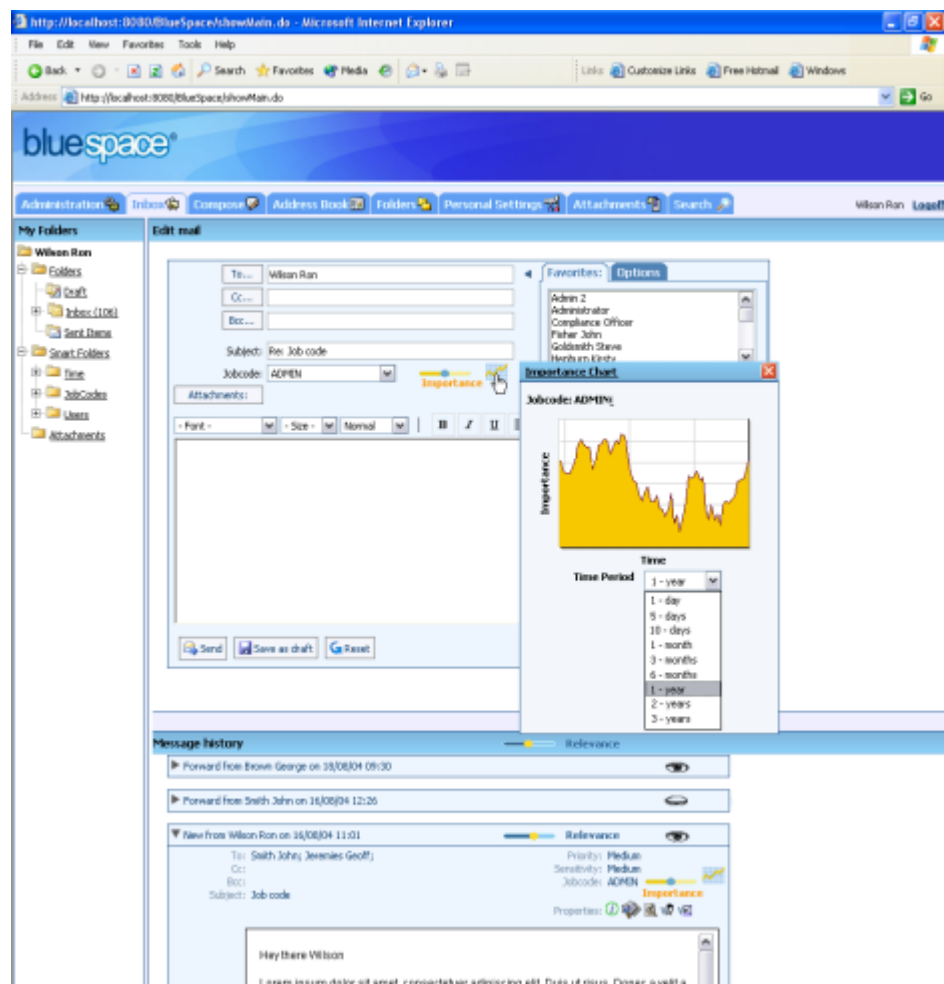


Figure 11 – Importance evolution

The evolution of importance over time is a critical concept in searching a set of TransMails, as discussed in detail in Section 7.3.

7.2.2 Folder view

The folder view is the way BlueSpace presents a TransMail folder listing to the user, showing select details of, say, 20 different TransMails at the same time. This is illustrated in Figure 12.

The screenshot shows the BlueSpace TransMail web interface. The browser address bar displays the URL: http://localhost:8080/blue-space-transmail/. The interface includes a sidebar with 'My Folders' and a main table listing email items. The table has columns for Relevance, Subject, Date, and other metadata. A relevance slider is visible for each row.

Relevance %	Subject	Date	Relevance %	First Read %	Last Read %	Sanitized %	Priority %
27/08/04 14:14	TransMail	27/08/04 14:15	27/08/04 14:15	27/08/04 14:15	Medium	Medium	
27/08/04 14:13	TransMail	27/08/04 14:13	27/08/04 14:13	27/08/04 14:13	Medium	Medium	
27/08/04 14:13	TransMail	27/08/04 14:13	27/08/04 14:13	27/08/04 14:13	Medium	Medium	
27/08/04 09:10	TransMail	27/08/04 09:11	27/08/04 09:11	27/08/04 09:11	Medium	Medium	
27/08/04 08:46	TransMail	27/08/04 12:38	27/08/04 12:38	27/08/04 12:38	Medium	Medium	
27/08/04 21:02	TransMail	27/08/04 14:11	27/08/04 14:11	27/08/04 14:11	Medium	Medium	
27/08/04 13:01	TransMail	27/08/04 13:04	27/08/04 13:04	27/08/04 13:04	Medium	Medium	

Figure 12 – BlueSpace TransMail folder view

The table in Figure 12 includes a column called relevance, which shows a relevance slider for every TransMail in the folder listing. As in the mail view in Section 7.2.1, the user can move the relevance slider to provide feedback to the application. In addition, the user can click the details icon which presents the same new box as that shown in Figure 10, indicating users' actions and importance indicators.

7.2.3 Address book

Figure 13 shows the BlueSpace address book. Each user (as well as each group and domain) is shown, and can be added to a user's favourites (those

they communicate with regularly).

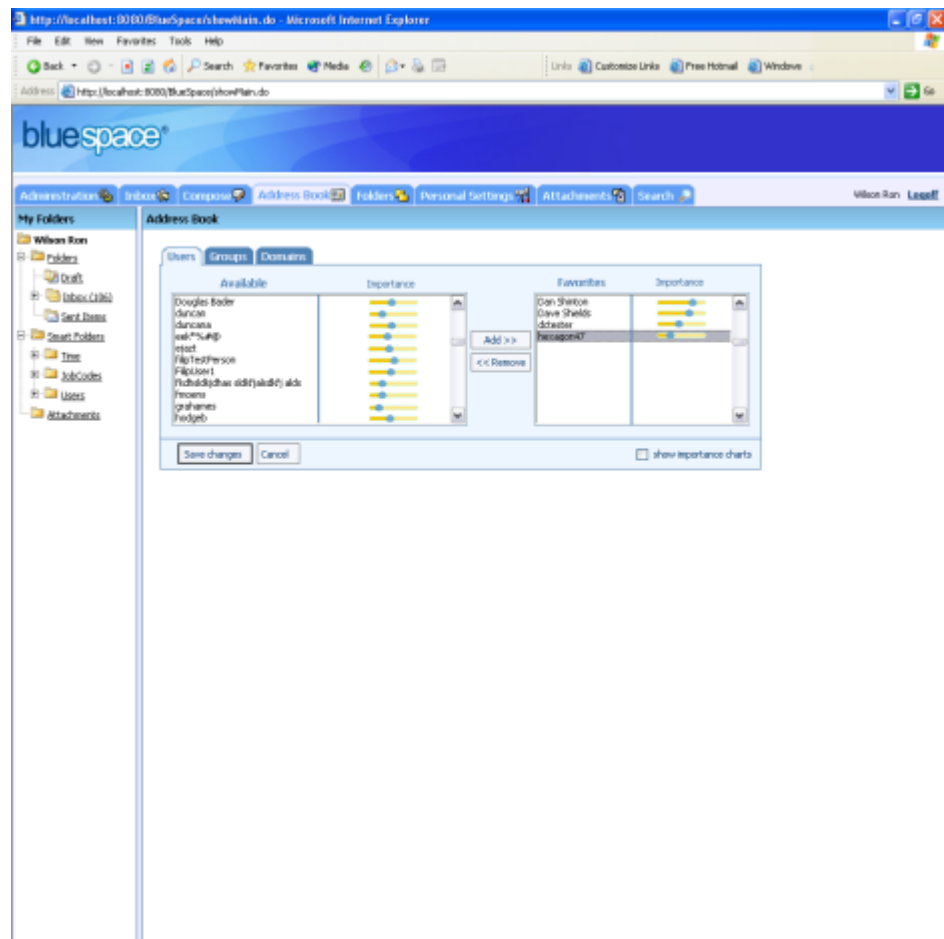


Figure 13 – BlueSpace address book

Figure 13 also shows an importance slider for each user, group or domain, and allows the user to directly interact with the importance values given to their contacts. Again, the user can either lock the value or provide it as a new start value from which the system equilibrates. Note that when user A adds user B to their favourites, the importance of user B with respect to user A is given a boost (in the algorithm).

7.2.4 Preferences

Using preferences the user can affect the different inputs into the relevance algorithm. This allows the user to tune the relevance algorithm to provide the content they are interested in. They can affect the relative strength of the different actions by users, as well as affecting the relative strength of the different entities with importance (Jobcodes, users etc). This is depicted in Figure 14.

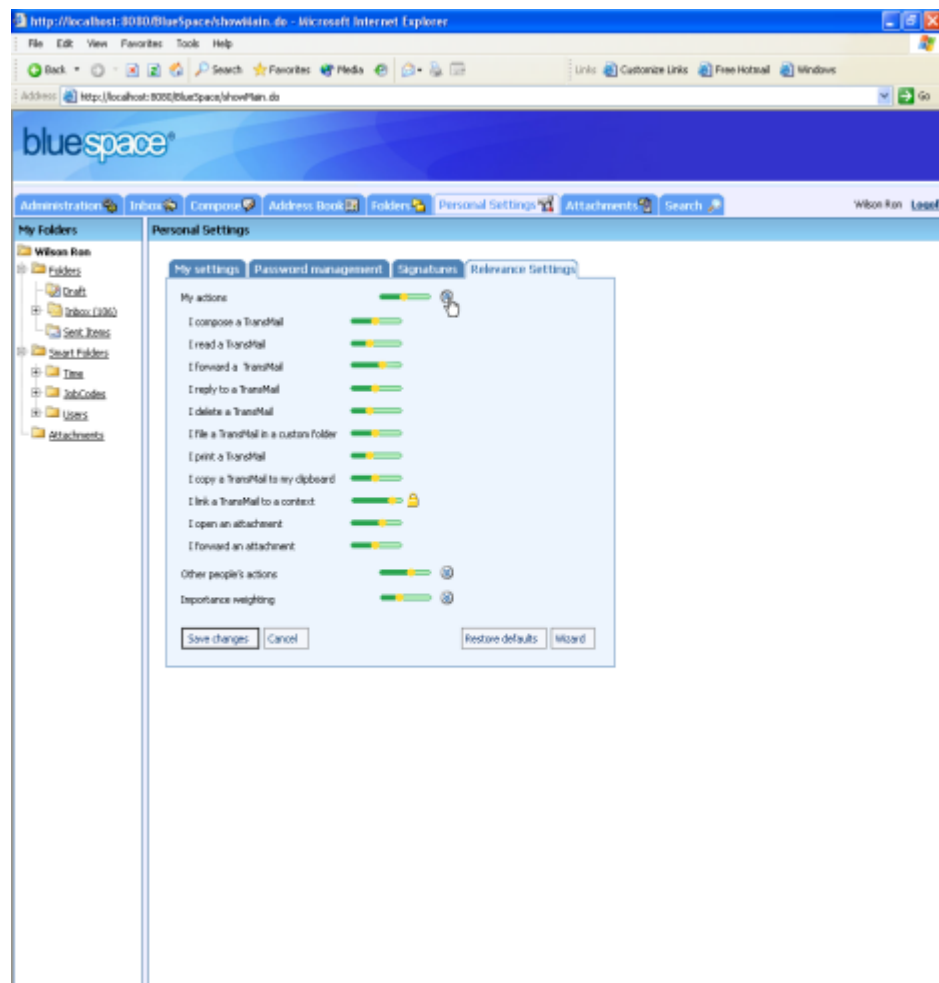


Figure 14 – Tuning the algorithm

The icons shown in Figure 14 are described in Table 3.

Icon	ALT Text	Action
	Change the significance of this action on relevance	Tuning slider icon allows the user to change the level of impact each action or entity has on the relevance algorithm.
	Significance locked	Allows the user to unlock the significance value, so it can vary again.
	Hide	Collapses a particular section of the UI.
	Show	Expands a particular section of the UI to show the respective actions or entities.

Table 3 – Algorithm tuning icons

In addition to varying the relative strengths of various inputs into the relevance algorithm, the user can also alter the importance of the different entities in the system, using an interface similar to that depicted in Figure 13 linked from Figure 14. The user can either lock the input at that value, or let it equilibrate over time – the slider for the impact of linking content to a context is locked in Figure 14.

The actions listed in Figure 14 can be customised, such as adding “I saved a local copy of a TransMail”. Such customisations can be done on implementation or by the system administrator.

Note that the mechanism of surveying users outlined in Section 7.2.1.2 provides the raw data that usually determines the weighting values. A given user's input partially determines both their own weighting values and (to a much smaller extent) the overall default weightings across the entire system. The reason for taking a content surveying approach to determining weightings is that weightings are a very esoteric factor, and it is very difficult for a typical user to determine these values directly.

In highly specialised systems, determining the weightings values can be done via a consulting exercise by the Systems Integrator (SI) implementing the system. The application has the ability to serve automated surveys to users via the weightings interface, in which users are 'walked through' various scenarios, and their responses determine the weighting values. This can be provided via the 'Wizard' button in Figure 14. The user can also restore the sliders back to their default values.

Another option for searching is to display the simple search capability, as discussed in Section 7.3, on every page of the application, allowing the user a familiar search function from wherever they are in the application. This choice appears in the "My settings" tab in Figure 14.

7.3 BlueSpace TransMail Search

BlueSpace TransMail search has been discussed extensively during this white paper, and it should be of little surprise that the search interface has far more flexibility and intelligence than any other mail search system. The search application can be divided into two areas – the search form and the search results.

7.3.1 Search form

The simple search form interface is presented in Figure 15.

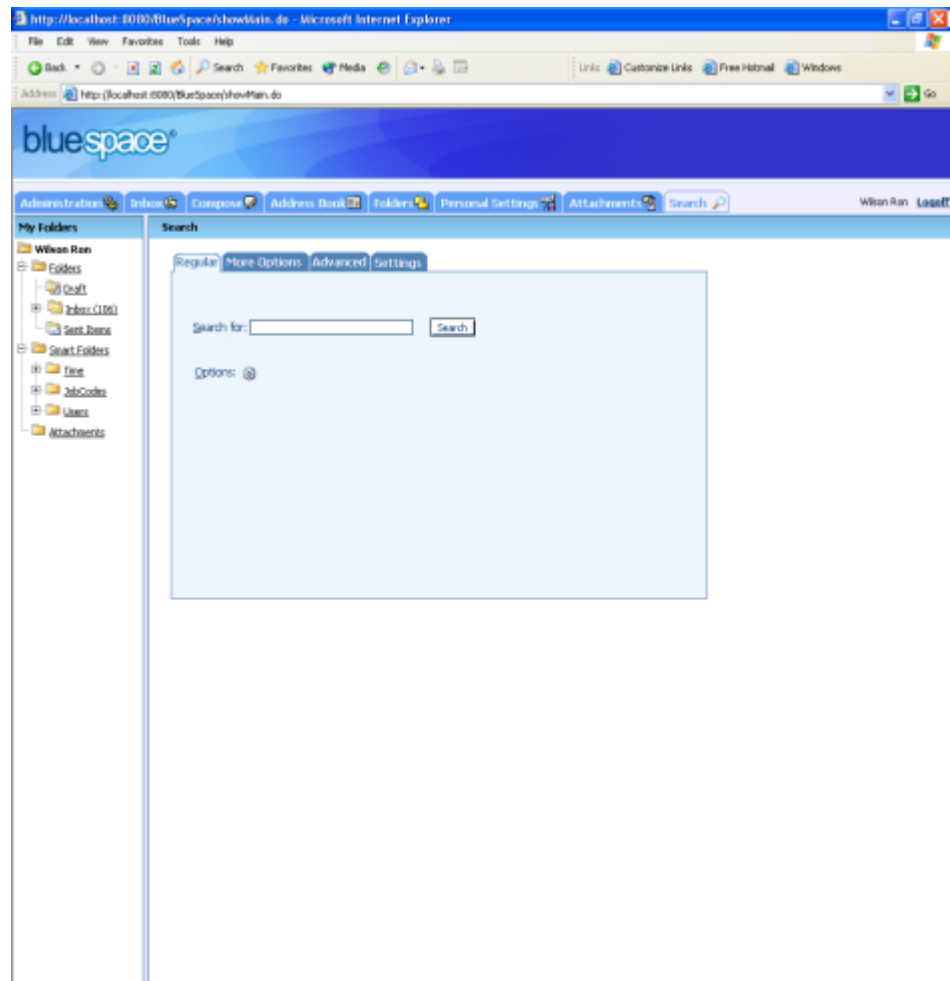


Figure 15 – Simple search interface

The tab shown is the simple search option, in which a user just enters any keywords they are interested in. The search algorithm then searches for this text in all possible criteria – users, groups, Jobcodes, message subjects, message bodies etc. This option provides an incredibly simple search algorithm for an average user.

If the user clicks the 'options' roll down in the search form interface as shown in Figure 15, the application then provides some additional search options, as shown in Figure 16. Note that the application 'remembers' when the user clicks the options roll down, and serves the search interface to them with the options already showing the next time the user goes to the search page.

The options roll down allows the user to specify which mail folders to look in, people involved, key words to search for, the date range and whether the

TransMail should have attachments.

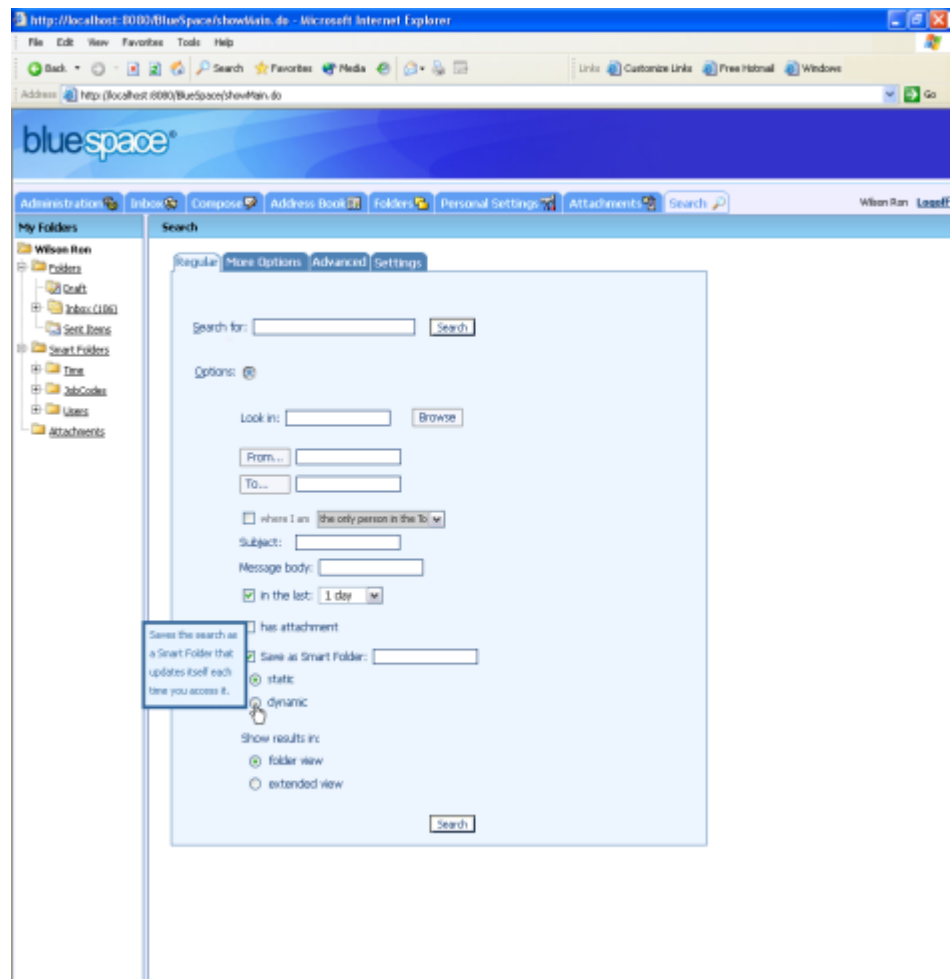


Figure 16 – Regular search interface

The search interface in Figure 16 provides the user with two pairs of options relating to what happens to search results. Each of these options has descriptive help in the form of a pop-up that appears when the user moves their mouse over the option.

A user can save any search they are doing as a Smart Folder that appears in a Custom folder in their Smart Folder tree. A static Smart Folder keeps the historical results of the search, so the user sees the same result set each time they click the Smart Folder. A dynamic Smart Folder carries out the search again each time the user clicks the Smart Folder. The user manages Smart Folders through the standard folder management interface. To edit any given Smart Folder, the user clicks the folder, and the application presents the folder's respective search query through the appropriate search interface.

The final decision for the user to make is whether they would like the folder view or extended view for their search results. These two views are discussed in Section 7.4.1.

The more options tab shown in Figure 16 provides another screen similar to that in Figure 16, but with more options for the user to choose through checkboxes, text fields, etc.

7.3.2 Advanced search

The advanced search interface is shown in Figure 17.

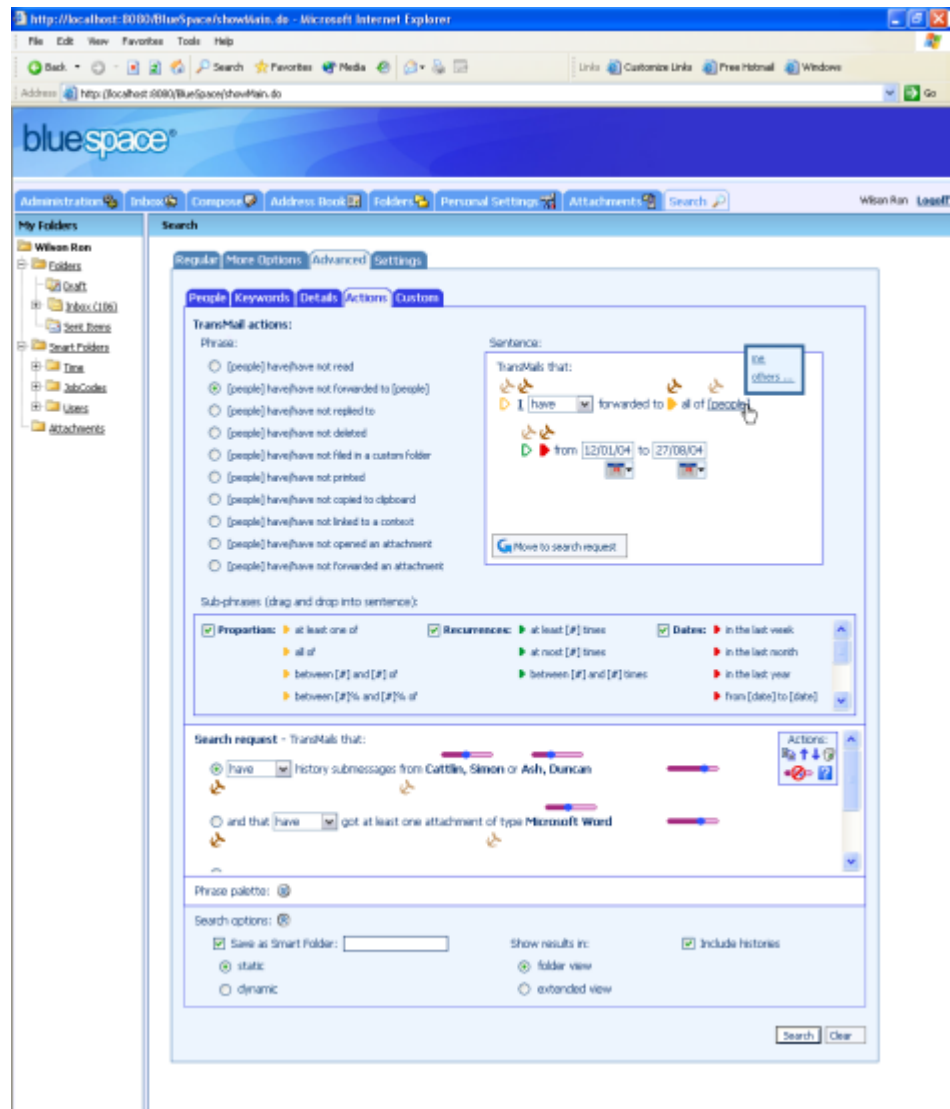


Figure 17 – Advanced search interface

Figure 17 may look rather complex, but the idea behind it is intended to be intuitive. The user can build search sentences, and then use them to do searches. The description of Figure 17 will be broken down into 5 sections:

1. Sentence types
2. TransMail actions
3. Search request box
4. Phrase palette
5. Search options

The icons shown in Figure 17 are described in Table 4.

Icon	ALT Text	Action
	Empty anchor	On click, shows a popup that indicates “You can move a phrase or sentence here by clicking a phrase anchor and dragging it here.”.
	Phrase anchor	Click and drag allows the user to move the phrase to an empty anchor.
	Empty subphrase	On click, shows a popup that indicates “You can move a subphrase here by clicking a phrase anchor and dragging it here.”.
	Subphrase	Click and drag allows the user to move the subphrase to an empty subphrase. (This icon then becomes empty.)
	Subphrase	Click and drag allows the user to move the subphrase to an empty subphrase.
	Calendar	Launches a calendar popup that allows the user to select the date.
	Change the weighting of this input to the search	Weighting slider icon allows the user to change the level of impact each action or entity has on the search request.
	Copy sentence	Creates a copy of the selected sentence.
	Move sentence up	Moves selected sentence up the ranking.
	Move sentence down	Moves selected sentence down the ranking.
	Delete sentence	Deletes selected sentence.
	Hide weighting sliders	This button hides the weighting sliders, no longer showing them in the search request box. On clicking, the icon toggles to a button without the red circle and line, that when clicked shows the weighting sliders again.
	Get help!	Launches a help and debugging tool.
	Hide	Collapses a particular section of the UI.
	Show	Expands a particular section of the UI to show the respective actions or entities.

Table 4 – Advanced search icons

7.3.2.1 Sentence types

There are five sentence types in the advanced search interface:

1. People – user, groups and domains that the mail is from and to.
2. Keywords – keywords in any field in the system, e.g. the subject or message body.
3. Details – details and properties of the mail, e.g. whether it has attachments or has been assigned a Jobcode.

4. Actions – user’s actions on mail content, such as reading or replying.
5. Custom – access to a library of other sentences.

These five sentence types are presented to the user through a tabbed interface in Figure 17, allowing the user to switch between the different types as they build up their search request.

7.3.2.2 *TransMail actions*

TransMail actions represent the largest area of screen coverage in Figure 17, because it is that sentence type that is shown in the interface. There are three subsections within this part of the view:

1. Phrase – the area in which the user selects the type of action they wish to use in their search request.
2. Sentence – the area in which the user builds up the sentence they wish to add to their search request.
3. Sub-phrase – the area from which the user can introduce modifications to their basic phrase.

In Figure 17, the action selected is that of a person forwarding a mail to other people (be they users, groups or domains). This phrase appears in the sentence box, allowing the user to ‘fill in the blanks’. In Figure 17, the user has already selected the first blank – that they themselves forwarded the mail. The user is in the process of filling in the second blank. By clicking the [people] field with their mouse, they have caused the application to spawn a pop up window. This window allows them to select the most request choice (themselves) or go to a modified view of their address book to select the people they wish.

The user can customize the phrase in the sentence box by adding in sub-phrases from the sub-phrase screen area. This part of the interface works a little like building bricks for kids. Each of the different sub-phrase types has a color coordinated arrow next to a given sub-phrase and a matching slot in the sentence box. The user can drag and drop sub-phrases into the sentence box, and then customize them as they wish. Figure 17 shows two sub-phrases have been added in – the “all of” sub-phrase and the “from [date] to [date]”. The date sub-phrase requires further user interaction to specify the dates, which the user has already carried out using the associated calendar buttons. Each calendar button spawns a DHTML calendar that allows the user to specify a date. Note that date formatting is a user preference determined globally across the application via the “Personal Settings” tab.

Once a user has built a search sentence that they wish to add to their search request, they can click the “Move to search request” button, which will add the sentence to the bottom of the list in the search request box. In addition, the user can use the anchor system. Any hard anchor (shown in dark brown in Figure 17) can be dragged to a light brown (empty) anchor slot, subject to certain rules regarding how phrases can be combined. The anchor system is common to the sentence area, the search request box and the phrase palette.

Note that in order to assist the user, there are various user popup help boxes that appear when the user moves their mouse over the different parts of the interface.

7.3.2.3 *Search request box*

The search request box is the screen area in which the user builds up their search criteria, drawing on sentences from each of the different sentence types. Each sentence is shown, along with anchors allowing each sentence to be extended with other sub-phrases. For example, the sentence, “have history submessages from Cattlin, Simon or Ash, Duncan” can be extended at the

“from” point to “have history submessages (from Cattlin, Simon or Ash, Duncan) or contain the keywords ‘sales proposition’”.

Each sentence (in both the search request box and the sentence box) can be changed between positive and negative logic. The user can change the “have” dropdown box to “have not”.

There are weightings for each sentence and each of its sentence criteria. These are shown by the purple sliders in Figure 17, and the user can move the positions of these sliders to affect the relative importance of the respective inputs in the search request. In the example shown in Figure 17, TransMails containing history submessages from Simon Cattlin would be given higher ranking than those from Duncan Ash in the search results listing. Also, the first sentence regarding history submessage authorship would be given slightly greater weighting than the requirement to have at least one attachment.

The final part of the search request box in Figure 17 is the “Actions” box. This contains icons that allow the user to manage the sentences that make up their query. The top row from left to right allows the user to copy a sentence (e.g. to search for two very similar sentences), move a sentence up the list, move a sentence down the list and delete a sentence.

The icon on the left in the second row allows the user to toggle the weighting sliders on and off. When the user turns the weighting sliders off, the ordering of the list is particularly important, as the system automatically attaches greater weighting to sentences further up the list. Note that the icon shown in Figure 17 has a red circle with a line through it indicating it is to turn the weighting sliders off. If the weighting sliders were hidden, the icon would just show the slider to indicate that it would toggle the weighting sliders back on again.

The final icon in the actions box is a “?” symbol that provides the user with help on using the advanced search interface. When clicked, the icon launches a debugging interface in a pop up that walks the user through debugging their sentence to ensure that it is consistent. This interactive debugging is highly useful in helping to ensure that the interface is usable by ordinary people, as well as being both sophisticated and powerful.

Note that if a user starts designing a search in the regular search interface shown in Figure 16, and then moves to the advanced search interface, their initial input is represented as a sentence in the advanced interface. The user can then further customise and extend it from there. This provides an easy first step for a user attempting to move from the regular search interface to the advanced one.

7.3.2.4 *Phrase palette*

The phrase palette was mentioned in Section 7.3.2.2. It provides a ‘dumping ground’ for the user to move phrase and sub-phrases about. It is shown rolled up in Figure 17, but the user can expand it by clicking the chevron.

The phrase palette is particularly useful to the user when transferring phrases from their custom sentence type to mix with other sentence types. The custom folder and phrase library is outlined in Section 7.3.3.

7.3.2.5 *Search options*

Many of the elements in the search options box are the same as those already introduced in Section 7.3.1. The user can save a search as a smart folder that appears in their Custom smart folder set. The smart folder can either be static, in which case it keeps the same result listing as when the search was first run, or dynamic, in which case the application reruns the query each time the user views the smart folder.

Search results can either be displayed in folder view or in extended view, as discussed in Section 7.4.1. An additional option for the user is whether to

include the historical submessages or not. Underneath the search options box lie the “Search” and “Clear” buttons.

7.3.3 Custom phrases

The custom sentence type provides access to a library of phrases and sub-phrases. This library is typically a centralised resource held server-side across the company. The user can type in phrase fragments and search for phrases that match their desired input – bridging the gap between natural language searching and logic-based searching. The corporation can also update this library with new phrases and sub-phrases as they see fit, mapping these to new logic elements at the search level.

Examples of additional phrases include:

- “that I wrote that became part of a tree linked to Jobcode X”
- “that I wrote and were read by Clarke, John”

Other phrases act to alter the search algorithm’s action, such as:

- “that will help me find someone’s name”
- “that will help me find a historical document”
- “that will help me find out about a project”

These phrases alter the search results in such a way as to assist the user in finding the type of information they are looking for.

7.3.4 Search settings

The user can interact with the settings for the search algorithm and result

ranking, as shown in Figure 18.

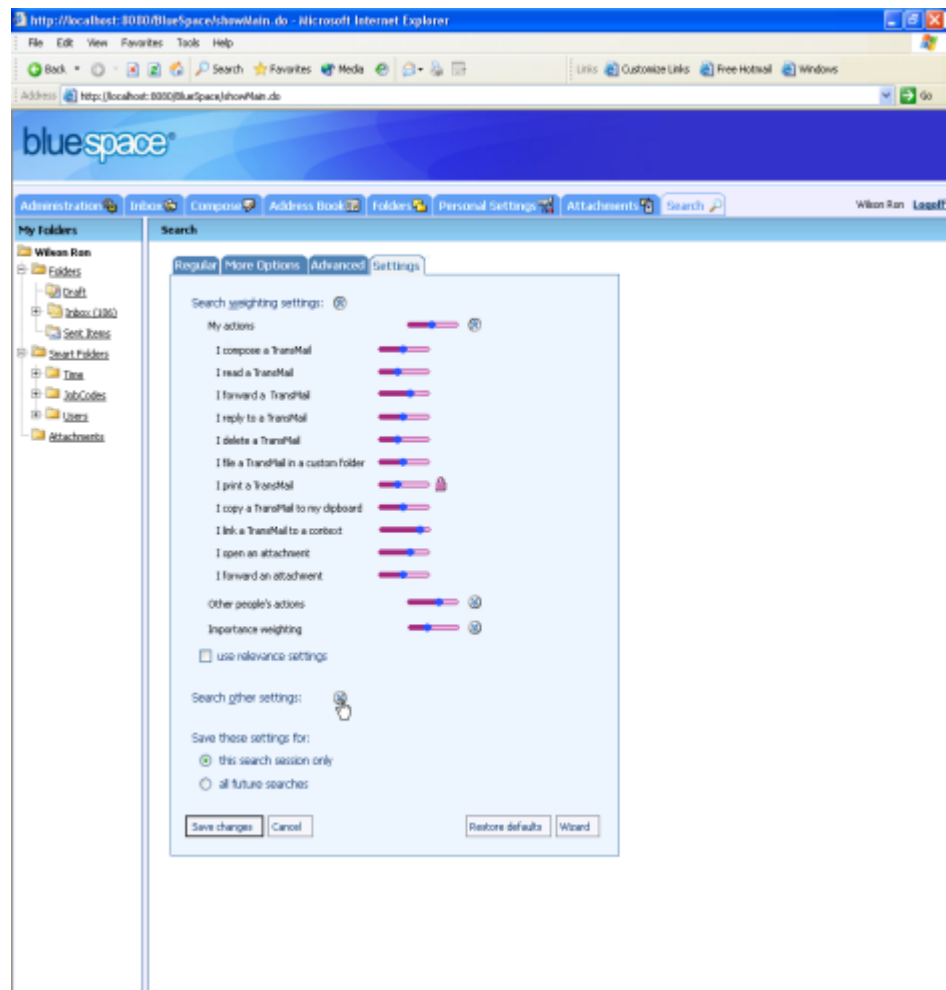


Figure 18 – Search settings

The icons shown in Figure 18 are described in Table 5.

Icon	ALT Text	Action
	Change the weighting of this action on search results	Weighting slider icon allows the user to change the level of impact each action or entity has on the search algorithm.
	Weighting locked	Allows the user to unlock the weighting value, so it can vary again.
	Hide	Collapses a particular section of the UI.
	Show	Expands a particular section of the UI to show the respective actions or entities.

Table 5 – Search settings icons

Figure 18 closely resembles Figure 14, as the user can change the relative strengths of the inputs to the search algorithm in just the same way as they can for the relevance algorithm. The user has the option of tying the weighting settings to be the same as their relevance settings.

The user has additional options they can change under “Search other settings”, including:

- The number of results per page in both search result views

- The information to include in the extended view
- The columns to include in the folder view

In addition, the user can add in new buttons that allow them to further alter the search results. These are accessed via a library similar to that described in Section 7.3.3. Examples of the functionality of these custom buttons includes:

- “Limit it to mail I read”
- “Show me mail read by more people”
- “Show me more recent content”
- “Let me change the weightings” (presents the user with a pop up to change the weightings of their search criteria)

A user can enforce any changes on either a session basis (during a particular log in session of searches) or for all future searches. This option is important, as it allows the user to increase the weighting towards specific criteria for a given search in which those criteria are important.

7.4 *BlueSpace TransMail Search Results*

7.4.1 Search results

Most email clients always return search results in a view resembling a mail folder. The use of this representation dates back to the origins of email, when it was designed by technologists for technologists, and reusing an existing graphical interface (the folder listing) was easier than designing another one.

The primary reason for the prevalence of the folder view for search results from mail is the wide adoption it received in email's early years rather than its suitability for the purpose. Today there is a surprising lack of innovation in the way search results are presented to end users, just as the search mechanisms and algorithms are archaic.

BlueSpace TransMail search has two views for search results:

- **Folder view:** A representation based on the way users are accustomed to viewing search results.
- **Extended view:** A representation more closely resembling the way users are accustomed to viewing web search results.

Providing both views and then allowing the user to choose which they prefer as they conduct their search ensures user satisfaction. The application ‘remembers’ the user's previous choice, and defaults to this view each time the user conducts a new search.

7.4.2 Folder view

The folder view of search results is shown in Figure 19.

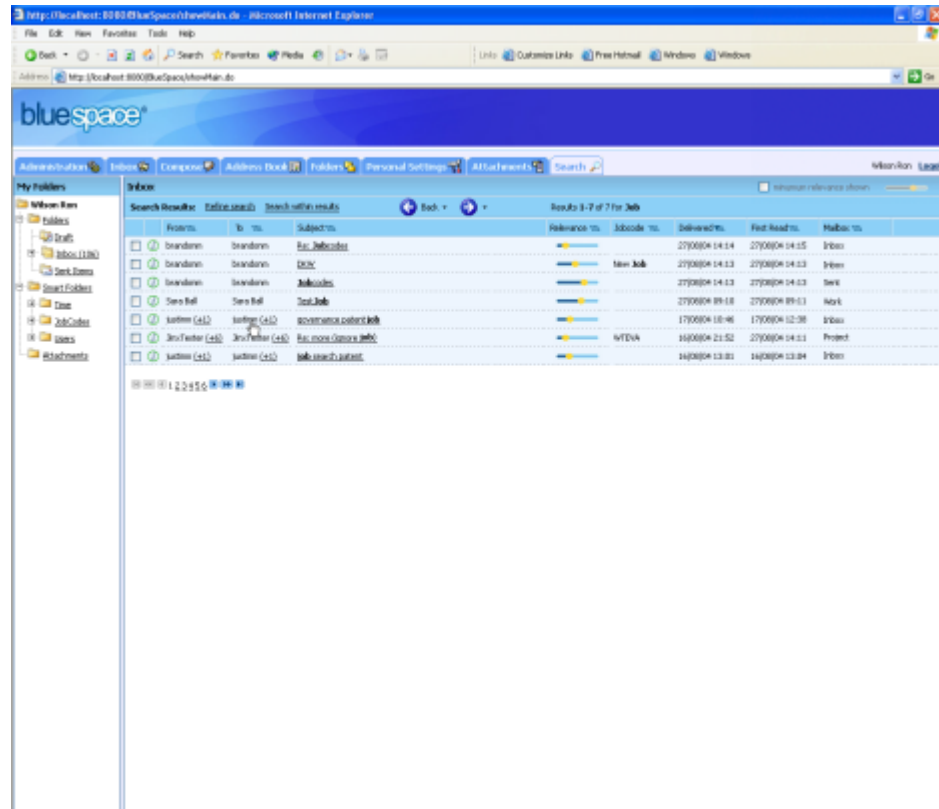


Figure 19 – Folder view search results

The icons shown in Figure 19 are described in Table 6.

Icon	ALT Text	Action
	Back to previous search	Takes the user back to the previous search they performed.
	Forward to next search	Takes the user forward to the next search they performed (only available when the user has already gone back in the search history).
	(Popup explaining value)	Relevance slider icon allows the user to change the relevance value for the content in question.
	More information	Spawns the popup shown in Figure 10.

Table 6 – Folder view icons

The view is intuitive, and designed to be such. The table indicates the author and recipients of each mail, the subject, any designated Jobcode, key dates (delivered and read) and the mailbox the TransMail resides in. In addition, the table includes an information icon that allows access to the same popup box as that shown in Figure 10, and relevance sliders for each TransMail returned by the search.

The row above the column headings includes some useful capabilities for the user. The user can refine the search and can also search within the results they have already been presented with. 'Back' and 'forward' buttons allow the user to move up and down their search history with a single click. The user

can sort each column ascending or descending to further customize the results view.

7.4.3 Extended view

The extended view of search results is shown in Figure 20.



Figure 20 – Extended view search results

The icons shown in Figure 20 are described in Table 7.

Icon	ALT Text	Action
	Back to previous search	Takes the user back to the previous search they performed.
	Forward to next search	Takes the user forward to the next search they performed (only available when the user has already gone back in the search history).
	Hide	Collapses a particular section of the UI.
	Show	Expands a particular section of the UI to show the respective actions or entities.
	Why was this found?	Launches a popup indicating why a given piece of content was shown as part of the search results.
	(Popup explaining value)	Relevance slider icon allows the user to change the relevance value for the content in question.
	More information	Spawns the popup shown in Figure 10.
	Find all copies of this submessage	Shows a custom Smart Folder (dynamic) with all TransMails that contain the given submessage.
	Find submessages like this	Shows a search result page in the default preference of the user (see Section 7.4), with submessages deemed to be similar to the given submessage.
	Visualise users	Shows a popup with a visualisation of the user interaction that took place to make this TransMail, similar to Figure 2.
	Visualise mail	Shows a popup with a visualisation of the mail interaction that led to this TransMail, similar to Figure 3.

Table 7 – Extended view icons

Figure 20 bears significant resemblance to web search results, as was indicated in Section 7.4.1. As in the folder view for search results shown in Figure 19, Figure 20 includes options to refine the search, search within results, and move backwards and forwards through search history.

The “News” content set is the results of searching a corporation’s knowledge base and intranet site – not TransMails. The first two items are links to external content that the search spider found on the intranet site, whilst the second two items have been written by other employees as part of the corporate knowledge base. The search results in BlueSpace are highly customisable, and additional content feeds can be placed into these top items. The corporation can assign content feeds to each user through the administrative interface, or can allow the user themselves to choose different content feeds. This functionality can be acquired as a module that acts as an extension to the BlueSpace Integrated Governance Management module.¹⁵

The next three items are shown rolled up, but the user can click the chevron to display each set of search hits. The content makes up these categories is shown below:

Contexts (4)

- Development
 - Denver
 - **Test**
 - UK
 - **UI test**
- Sales
 - TAM resources
 - **ROI tests**
 - **Deployment testing**

Jobcodes (2)

- Capita deployment
 - Capita **testing** process
- DKW deployment
 - DKW **testing** process

People (9)

- Users
 - John **Test**erea
 - **Test**in Cozonack
 - Suzanne **Etest**e
- Domains
 - Denver **Test**
 - UK **Test**
- Groups
 - Capita **testers**
 - DKW **testers**
 - UI **testers**
 - Volume **testing** group

Each of these sets constitutes a set of entities in BlueSpace, and the occurrences of the search term within them. The number shown in brackets denotes the number of search hits in each category.

Below these categories in Figure 20, the search results list hits within TransMail content. Each search hit highlights the content that caused the item to appear in the search results, with the following display schema:

- The subject of the submessage
- A magnifying glass containing a question mark
- The relevance of the submessage
- The body text showing the instances of the search term
- The people involved – sender and recipients
- A set of property icons identical to those in Figure 8

- The Jobcode or attachment name if they are part of the search hit
- The date the mail was received and read
- The mail folder in which the respective TransMail appears

Clicking a user (or group or domain) name causes the application to display the respective smart folder of the searcher's interactions with that population entity. Clicking a Jobcode, context or other entity has a similar effect.

Clicking the magnifying glass with a question mark inside spawns a pop up window indicating to the user how the result matches their query. This allows the user a quick way to understand why the result would be of relevance to their query, and how the application arrived at that decision. It also allows the user to provide feedback to the application (and thus algorithm), by indicating whether the result was what they were looking for.

Note that the relevance slider for the submessage with the subject, "Don't bother reading – just a test message" is quite low given that this is the first page of the search results. This highlights the fact that the relevance sliders are not respective to the search being conducted – they are absolute based on the output of the relevance algorithm. Relevance in search, described in Section 5.1, is used to determine the ranking for the search results (though may also be displayed as a user preference). This test submessage was deemed to have low relevance as few people read it, given it was a test and not real content. However, the fact that the body of the submessage contains many instances of the word "test" helped drive the submessage up ranking of the search results given the query was for the word "test".

The search result, "Don't bother reading – just a test message" does not show the instances of the search term in the message body. This is because the submessage is unread, and the governance engine in BlueSpace does not allow partial reading of a piece of content in a user's incoming TransMails. The user must click the submessage to read it if they wish to view the content.

There are two options for search term highlighting. That shown in Figure 20 display the search term in bold to help the user identify the terms. Another option (at the bottom of Figure 20) is to highlight the search terms – this is particularly useful for searches for multiple terms, as it highlights each term in a different colour.

The language of this section has indicated that each search hit is in fact a submessage, not a TransMail per se. It is up to the user to decide whether they wish to only search 'current submessages' as part of TransMails, or whether they wish to include historical submessages. This option was discussed in Section 7.3.2.5.

The page numbers at the bottom of the screen in Figure 20 are outside the iFrame containing the search results, and allow the user to move between pages of results.

7.5 Summary

This section has overviewed the search interface in BlueSpace with respect to the end user – how they see it. It contains a number of novel features that improve the user's search experience. By introducing real-time interaction with a relevance algorithm, mail content can be leveraged as the knowledge base it has become rather than just a 'content skip' in which no-one can ever find what they're looking for.

8 Blogs

8.1 Introduction

Blogs are an upcoming area for communication on the internet, and follow the 'smart pull' as opposed to push model. In essence, a user subscribes to a blog like a newsletter, and this data is then automatically pulled into their user interface by the application client (e.g. Microsoft Outlook).

There are a number of advantages for blogs over email for certain types of communication, such as:

- They are automatically structured by author or other criteria.
- They do not block up the inbox like newsletters.

In many ways, the popularity of blogs can be seen as a user backlash against mail overload.

8.2 Relevance In Blogs

While this white paper has focused on applying the BlueSpace relevance system to a mail environment, the same system can be applied to blogs. The actions of users can be recorded, the relevance of given blog listings calculated, and the importance of the various entities involved inferred.

BlueSpace relevance in blogs has been applied to both the BlueSpace blogs model and also to other open-standards blogs and blog interfaces using a plug-in approach similar to that described in Section 6.4 and illustrated in Figure 7.

For brevity, relevance in blogs is not described in full, as it is highly analogous to that in mail environments.

9 Algorithms

9.1 Brain Dump

We need to compare to Google's page rank technology, and there's a great study of it here: <http://www.webworkshop.net/pagerank.html>

Other links:

<http://ausweb.scu.edu.au/aw2k/papers/lowe/paper.html>

Need an expert in this area.

Need to discuss the way BlueSpace does (or will) split up content into a query store and a transactional store.

Calculate relevance for content as it comes in, then change it each time a user does an action, then weight it for people relevant to me, then recalculate importance scores for people on say nightly basis.

May be that due to outliers and random behaviour, if one person leaves a mail open for 30 minutes it's less than 10 times relevant than if ten people do. Need a way to deal with outliers – but not likely to get more than one.

If you forward to your boss and that's more important than one you sent to the office secretary, so you need to run the algorithm iteratively. This is because if the boss is of higher importance the secretary, then forwarding a message to the boss from Bob increases Bob's importance more than if you forwarded it to the secretary.

Importance of content can be done in realtime, importance of people can be reassessed each night.

Taking lag time – so longer you wait before reading, the less important.

Most others deleting mail – might want to weight the users again.

Different people can have different relevance ratings for different respective users – my boss matters more to me than they do to someone else in a different department. So need an n by n matrix.

Each user can affect the relevance scale for any other user as part of a learning cycle / feedback loop.

Big difference compared to normal search engines – each user gets different results.

Relevance scores need to drop off over time – so data that's two years old should have less impact than stuff that's 2 months old.

Relevance actually varies over time – during a particular date period a user may be much more relevant to me due to the project I was working on with them. This clearly reflects search, as relevance could be based on today's relevance values or the relevance value relative to the date the content was written.

Searches should go back into feedback loop – potentially has greater impact on future searches than on messages I'm interested in now, as relevance criteria changed over time. What you're interested in when searching may be very different to what you're interested in when reading – Dilbert cartoons may be the highlight of your day, but you never search back to find particular ones.

Multi-dimensional – if I email Andrew cc Yul and he always keeps it, then email Andrew cc Dan and he always deletes it, then different relevance.

Need to think about common themes in searches – how do you delve into this. Have all the data in a structured way, but how do you recognise themes and related themes through quasi-structured mail content.

Note that in the TransMail view, the TransMail relevance should be some algorithm summing the relevance of the submessages (both current and history) though the relevance components should drop off over time. So the further back in the history you go, the less effect that submessage has on the TransMail's overall relevance.

In order to assist the algorithm limiting the size of data it needs to pull in conducting the search, the administrator could reduce the initial data set. As an example, an administrator can determine that only the last 3 months of data are searched first, and the user is given a button saying "Expand search" to allow them to include older content.

From an algorithm perspective, the application can search first through content with a relevance over a given threshold value, which reduces the computational load and increases response times for building the first pages of search results. Also, typically a user will find what they are looking for in the first couple of pages of results, thus searching through to low relevance content is an unnecessary performance drain.

Note that the time spent reading a given TransMail may also be indicative of relevance, and could be built into the relevance algorithm.

We also need to build in the idea of a federated weightings model, in which corporations provide the derived values for weightings of different actions by web services to a central pool. This increases the sample size, and means that new systems don't start from random values – we build 'understanding' of users' action weightings across corporations. Later, this could also be classified according to different verticals, as health care weightings might be different to manufacturing or military systems.

9.2 Initial Maths

NOTE THAT THIS INITIAL MATHS HAS BEEN DRAFTED WITH A MATHEMATICIAN, BUT IS IN NEED OF SIGNIFICANT CHANGES. IT IS INCLUDED HERE FOR COMPLETION IN THIS CURRENT DRAFT.

9.2.1 Introduction

There are three lists of attributes which could feasibly influence the Transmail's relevance to a receiver user. The three lists are groupings according to mail actions relevance, collaborative relevance and defined importance factors, which I shall label A , B and C respectively. The most obvious starting point is to work with indicator functions, $J(\cdot)$, which are simply "on" or "off", depending on whether the attribute is favourable or not. This is fine for attributes such as:

"If a user receives a TransMail and replies to it, then the content of the original TransMail is likely to be of significant relevance",

but not necessarily for attributes such as:

"If a user has read a given TransMail many times, then it is likely to be of significant relevance to them.",

where a score function may be more appropriate, with a higher score dependent on the number of times opened or read. However this causes problems of relative weighting amongst attributes, and in any case we can create an indicator function by defining a threshold value of number of times opened or read. The third list is different as it deals with importances (see later), and thus must have a continuous scale of values. These importance

functions are called $I(\cdot)$. A working formula for the relevance, $R(T_M)$, of a given transmail, T_M is then:

$$R(T_M) = \frac{1}{3} \left(\frac{1}{W_A} \sum_{j=1}^{n_A} w_j^A J(A_j) + \frac{1}{W_B} \sum_{j=1}^{n_B} w_j^B J(B_j) + \frac{1}{W_C} \sum_{j=1}^{n_C} w_j^C I(C_j) \right)$$

where n_A , n_B and n_C are the number of attributes in each list, w^A , w^B and w^C are weightings associated with each attribute to allow for the possibility that some attributes are deemed of greater intrinsic relevance than others, and W^A , W^B and W^C are the sum of the weightings in each of the three lists. Of course if the weightings are all unity in a list then $W = n$ for that list. This formula assumes that the relative maximum weighting of each list is equal to 1, a sensible option; it may also be sensible to weight each list in the ratio $W_A : W_B : W_C$, in which case the formula would be:

$$R(T_M) = \frac{1}{W} \left(\sum_{j=1}^{n_A} w_j^A J(A_j) + \sum_{j=1}^{n_B} w_j^B J(B_j) + \sum_{j=1}^{n_C} w_j^C I(C_j) \right)$$

where:

$$W = \sum_{j=1}^{n_A} w_j^A + \sum_{j=1}^{n_B} w_j^B + \sum_{j=1}^{n_C} w_j^C$$

the sum of the weight within and across the lists.

9.2.2 Importance

The third list is special in that it introduces the interactive aspect of relevance between users through entity importance, whereas the first two are purely independent in their individual factor contributions to relevance. This is because the intrinsic nature of these factors for mail actions by the user or by other users other than the user in question will affect the relevance of the Transmail, but the relevance of the Transmail will not feedback into the quantification of these factors. They either happen or they don't – the fact that the relevance may dictate whether these actions are engaged is irrelevant here, as we are merely revising relevance in the light of **past** actions. However, the importance factors are different. Here the values of these factors are affected by the email traffic that has occurred. For example the importance of a user is affected by the receipt of Transmails from other users – his/her importance perturbs or is shared by the importance of the other users. This interactive process must be taken into account before the importance can be fed into the equations (1) or (2) in a search. Similar arguments will apply for other importance factors, so we will assume that we can focus on one such importance factor, say the user importance factor.

To estimate the importance of a user the following model can be considered.

1. There is, for user i , an intrinsic immutable importance, I_i^0 , which is set globally and presumably takes into account levels of responsibility within the organisation. This intrinsic value should ensure that the mechanism doesn't result in total uniformity of user importance.
2. If a Transmail is sent from user i to user j , the importance of user j is amended slightly such that if user i is of higher importance than j

then the importance of j is raised; however if the importance of i is lower than j then the importance of j is lowered. There is in essence a sharing of importance from the sender to the receiver. The importance of the sender is unaltered by the Transmail, but of course his/her importance **relative** to the receiver has been altered, such that the difference between them has been reduced.

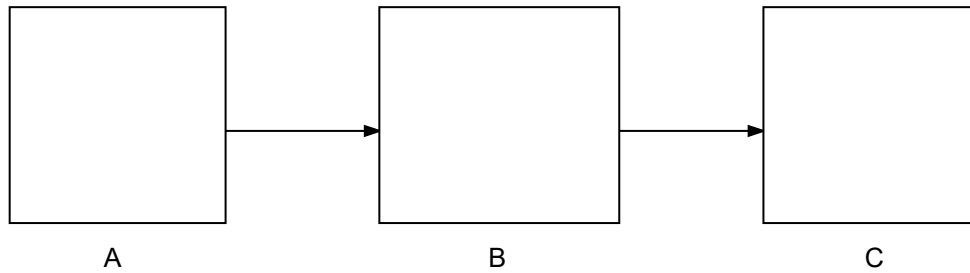
3. The “sharing parameter”, α , dictates the degree of this sharing, with $0 < \alpha < 1$. In practice α will probably high, say 0.9 or higher. There is also a parameter β to allow for the restraining to the preset level, I_j^0 .

With this, an equation to update user j 's importance at the k th update from Transmail's received since the last update, $k-1$ is:

$$I_j^{(k)} = \beta I_j^0 + (1 - \alpha - \beta) I_j^{k-1} + \frac{\alpha}{m} \sum_{i=1}^m I_i^{k-1} / C_i$$

where C_i is the number of recipients of the i th email. Note in the summation the same user may appear more than once. Note unless we divide the third term in the formula above by m , we will always increase the importance of a user by receipt of Transmails. This should not, of course, be the case. Furthermore if no emails are received the importance is slightly reduced; this is feasible as it indicates a nonparticipatory aspect of the user.

9.2.3 Example



User A has sent a Transmail to B who has sent a Transmail to C. Here:

$$I_A^k = \beta I_A^0 + (1 - \alpha - \beta) I_A^{k-1}$$

$$I_B^k = \beta I_B^0 + (1 - \alpha - \beta) I_B^{k-1} + \alpha I_A^{k-1}$$

$$I_C^k = \beta I_C^0 + (1 - \alpha - \beta) I_C^{k-1} + \alpha I_B^{k-1}$$

This system of equations will be solved iteratively in what is essentially a Gauss-Seidel algorithm procedure until convergence to a sensible precision. This procedure will then be repeated every regular time period, say daily. In this example, the more important A is, the more important B becomes, which then increases C's importance more than if B had not received the email from A.

The importance model above allows for relative importance for B with respect to A, as the adjustment resulting from a Transmail between A and B depends on the relative values for I_A and I_B .

9.1.4 Time-dependent relevance

Returning to the main framework equation for relevance (for example equation (1)), we would like a received Transmail's relevance to generally a time-dependency. We should therefore premultiply the formula in equation (1) by a sensible function of time, $G(t)$:

$$R(T_M) = \frac{G(t)}{3} \left(\frac{1}{W_A} \sum_{j=1}^{n_A} w_j^A I(A_j) + \frac{1}{W_B} \sum_{j=1}^{n_B} w_j^B I(B_j) + \frac{1}{W_C} \sum_{j=1}^{n_C} w_j^C I(C_j) \right)$$

For example $G(t)$ might reflect relevance of a project up to a deadline time, t_0 , beyond which the relevance begins to decay:

$$G(t) = \begin{cases} 1 & t \leq t_0 \\ e^{-\lambda(t-t_0)} & t > t_0 \end{cases}$$

where λ is a decay constant (high for rapid decay, low for slow decay).

9.1.5 Notes

1. These formulae are additive in their nature between the three lists, but for importance is interactive within. A further degree of sophistication might be to include degrees of effect, say on user action which is dependent on the importance of the user or other entities. These are then interactions between the actions and the importances.
2. There is no effect on the sender's importance of the Transmail in this model. We could model alternatively with the adjustment to importance being made to the sender rather than the receiver, or both.
3. The model for importance is reminiscent of that used for the PageRank system by Google, but is not exactly the same. The same iterative procedure, which will employ the basic Gauss-Seidel algorithm is employed however.

10 Conclusion

Email is now the most voluminous communication medium on the planet. Business users are drowning in a 'torrent' of content. While in part much of this may be external Spam, much of this email in large corporations is internal Spam through dysfunctional email use and the simplicity of the Cc: function in email applications. A typical business user can receive over two hundred 'real' emails (excluding external Spam) each day.

BlueSpace Mail Relevance is designed to alleviate the email pressure on end users – improving corporate efficiency and helping to re-establish 'quality of life' for end users.

BlueSpace Mail Relevance watches how users interact with content and implicitly other users and entities in the system, and takes each user's behaviour as the primary input into the relevance algorithm. The attraction of this approach is that it collects vast amounts of statistical input data to the algorithm, but does not require additional user time. BlueSpace Mail Relevance watches and learns.

BlueSpace Mail Relevance may be applied at a single user level only – based on a user's own actions. It can also be applied across an entire corporation in a collaborative manner, which greatly increases its accuracy by providing more contextual data.

The mathematics and computing behind BlueSpace Mail Relevance is certainly intricate and involved, but all of this complexity is hidden from the end user.

BlueSpace Mail Relevance will set a new standard for mail content handling in the global communication marketplace. In a business world in which everything is getting faster all the time, BlueSpace Mail Relevance will ultimately act as the 'swimming coach' and 'flotation aid' to corporate managers around the world.

11 Glossary

This document relies on the following term definitions:

- **Business Records Management (BRM)** – the market in which BlueSpace positions itself, abstracting content from communications that constitute business records, and managing this content in strict compliance with a company's policy.
- **TransMail** – the unit of mail in BlueSpace, and thus the directly analogous 'envelope' to email for users. A TransMail is made of a current submessage (which determines the TransMail subject, sender and recipients), and one or more history submessages.
- **Submessage** – a textual content fragment, authored by a user or computer, referenced by TransMails. Note that a submessage includes or may consist entirely of form components (e.g. check boxes and dropdown selection boxes).
- **History submessage** – a submessage that forms part of the history portion of a TransMail. Any TransMail sent as a reply or forward will have one or more history submessages.
- **Future submessage** – once a TransMail containing a given current submessage, A, has been sent, it may later be forwarded or replied to. Any submessage that becomes the current submessage of one of these later TransMails is a future submessage with respect to submessage A.
- **Attachment** – a file fragment, attached by a user or computer, referenced by TransMails.
- **Population Structure** – consists of all users and describes how those users belong to domains and groups.
- **Entity** – a collective name given to any actor or object in the system, i.e. a user, domain, group, context, Jobcode etc.
- **Domains** – sets of users. They are hierarchical in that domains may have sub-domains. These sub-domains default to having the same properties as the parent (though each property can be specifically changed). Domains are mutually exclusive.
- **Groups** – sets of users that span domains, such that users in different domains may belong to the same group. Groups are not hierarchical and are non-mutually exclusive, i.e. can overlap.
- **Relevance** – an indication of how likely it is that a given piece of content will be of interest to a certain user. Relevance can be applied to both content viewing and content searching, and is the result of an algorithm.
- **Importance** – a measure of the level of significance an entity has to another. For example, if John always reads and replies to mails from Sally, then Sally is important to John. Importance is an input to relevance. Note that a given entity will have lots of importance values, as these values are all relative to other entities.
- **Actions** – users' actions determine both relevance and importance values. Both direct user's actions and collaborative actions (actions of other users on the same content) are taken into account.
- **Slider** – a user interface feature that allows a user to view relevance and importance values similar to the bars of a graphic equalizer in a sounds system. The user can click and drag the slider's level to

provide feedback to the system, similar to changing the level of a bar on a graphic equalizer.

- **Importance chart** – a visual representation of the way the importance of a given entity has evolved over time.
- **Compliance Policy** – an overarching term defining all aspects of compliance in a BlueSpace system.
- **Rule Set** – a collection of rules that govern the actions that users can carry out in the BlueSpace system.
- **Rule Type** – rule types are categorised as follows: Communication Rules, Readership Rules, Action Rules, Retention Rules and Search Rules.
- **Rule Categories** – divides a Rule Set by rule type.
- **Standard Rule Set** – the Rule Set applied to messages if no Jobcode is specified – in other words, the default rules.
- **Communication Rules** – determines how users, groups and domains can communicate with other users, groups and domains.
- **Readership Rules** – determines which users, groups and domains can read a given piece of content after it has been written during its entire lifecycle (including mail forwards, etc).
- **Action Rules** – determines whether a given piece of content can be forwarded, copied to clipboard, saved, printed, replied to all etc. This Rule Type
- **Retention Rules** – determines for how long a given piece of content should continue to be retained.
- **Search Rules** – determines which content users can view when searching the entire message database.
- **Jobcode** – a way of linking any given submessage with the rules that affect the way that it is handled, allowing exceptions to the default behaviour described by the Standard Rule Set.
- **Compliance Administrator** – a ‘super-user’ who creates Compliance Policies in the BlueSpace system. Note this is a role (or privilege) given to a user, as are the other administrator roles listed below.
- **Administrator** – a ‘super-user’ who creates and manages domains, groups and users. Note this is a role given to a user, and a user may have more than one role given to them.
- **Auditor** – a ‘super-user’ who can view a log that indicates every action performed in the BlueSpace system.
- **Compliance Log** – a log that records any event that, if carried out, would contravene the Compliance Policy.
- **Audit Log** – a log that records every significant action that occurs in a BlueSpace system.

-
- ¹ [Talking Headers](#), Katie Hafner and Matthew Lyon, The Washington Post Magazine, 4th August 1996.
- ² [Legal Lottery](#), Melanie Liew, Computerworld, 28th April 2004
- ³ [Talking Headers](#), Katie Hafner and Matthew Lyon, The Washington Post Magazine, 4th August 1996.
- ⁴ [Simple Mail Transfer Protocol](#), Johnathan Postel, RFC 821, August 1982
- ⁵ [Is banning e-mail a viable way for IT to avoid being crippled by storage costs?](#), Julia Vowler, ComputerWeekly, 16th October 2003
- ⁶ [Email management too important for CIOs](#), Rachel Fielding, Network News, 11th February 2003
- ⁷ [Legal Lottery](#), Melanie Liew, Computerworld, 28th April 2004
- ⁸ *Relational Messaging White Paper*, BlueSpace Group Ltd, 2nd July 2002
- ⁹ *BlueSpace Discovery Service*, BlueSpace Group Ltd, 3rd December 2003
- ¹⁰ *BlueSpace Email Structuring*, BlueSpace Group Ltd, August 2004
- ¹¹ *BlueSpace Integrated Governance Management*, BlueSpace Group Ltd, 15th May 2004
- ¹² *BlueSpace Integrated Governance Management*, BlueSpace Group Ltd, 15th May 2004
- ¹³ *BlueSpace Email Structuring*, BlueSpace Group Ltd, August 2004
- ¹⁴ *BlueSpace Integrated Governance Management*, BlueSpace Group Ltd, 15th May 2004
- ¹⁵ *BlueSpace Integrated Governance Management*, BlueSpace Group Ltd, 15th May 2004

EXHIBIT 2006

metadata slip sheet

BegDoc	REV0000094
BegAttach	REV0000094
File Name	BlueSpace Mail Relevance.doc
LastCreateOn	4/18/2005 2:56 PM
LastModifiedDate	6/7/2005 4:16 PM
Custodian	Marston, Justin
Author	.
SortDate	6/7/2005 4:16 PM