Information Access Efficiency: a Measure and a Case Study

Shijian Lu and Cécile Paris

CSIRO ICT Centre
Locked Bag 17,
North Ryde NSW 1670 Australia

(Shijian.lu, Cecile.paris)@csiro.au

Abstract *One of the advantages we claim for information synthesis and aggregation is that it results in more efficient information access for end users, especially when the relevant information comes from multiple heterogeneous data sources. Although that claim is plausible, it has not been verified by any qualitative studies. It is even unclear how one would quantify the efficiency of information access. In this paper, we propose a measure and then report on a study to identify the information access efficiency gain of a potential application involving information synthesis and aggregation.*

Keywords information access efficiency, information aggregation, information access time and speed, information relevance.

1. Introduction

In the age of information, the competitive edge of an organisation is often defined not by how much information it possesses, but by how efficiently it can be accessed and how it will empower its staff to make good decisions quickly by accessing relevant information. What is relevant information is highly context sensitive. It could depend on the nature of the organisation, users' functional role or task.

In CSIRO, we are developing an information delivery platform (called the Myriad Platform) [5]. It enables the development of applications which automatically generate tailored documents, aggregating and synthetising information from a number of heterogeneous data sources. The effectiveness of tailoring has often been reported in the literature (e.g., in tutoring applications [3] or in medical applications [1]). But there has been no study as to the effectiveness of aggregation in supporting the information seeking task, independently of the tailoring. One of the advantages we claim for information synthesis and aggregation is that it is faster for end users to access relevant information, which could come from multiple heterogeneous data sources. Although intuitive and plausible, this claim has not been verified to our knowledge by any quantitative studies, performed independently of the tailoring. It is even unclear how one would quantify information access efficiency to be able to compare various approaches to accessing information. In this paper, we define a way to measure information efficiency, grounding a qualitative description on some measurable physical entities. We also present a specific study aiming at validating our claim that information aggregation and synthesis improves information access efficiency.

Our investigation is undertaken in the context of a specific potential application for presenting an end user with aggregated information. This application is set within a research organisation, and the scenario is as follows: in such an organisation, staff members often need to know about each other to a certain extent in order to effectively collaborate on various matters. They typically require several pieces of information about a fellow staff member. Currently, different information about a staff member is scattered in various data sources and needs to be accessed via different applications or access points. Our hypothetical application is a staff information web site where information aggregated from various data sources is immediately provided. This is a very simple scenario. Yet it gives us a concrete scenario allowing us to investigate issues around information efficiency: how much efficiency in information access would we gain by developing this hypothetical application?

To start with, then, we first identified what information staff are interested in when looking up a fellow staff member. Based on this information, we recorded exactly what it takes to find those information items within the existing organisation infrastructure, asking questions such as: how many applications/tools are needed? And what kinds of interactions are required? We mapped these interaction types to interaction time, and proposed some metrics to quantify the time required to obtain all this information using the current tools and an information efficiency measure. We then postulated that our hypothetical application would be able to aggregate all the information and present it in one virtual page to the information seeker. We compared the information access measures for the current situation and the hypothetical application to derive a measure of information access efficiency gain.

This paper is organised as follows. We first describe how to identify what information is relevant in the application context. Our approach for measuring information access efficiency is then elaborated. Finally, we conclude by discussing some implications of our study.

Proceedings of the 11th Australasian Document Computing Symposium, Brisbane, Australia, December 11, 2006.

2. Gathering information about fellow staff members

In the first part of our study, we used questionnaires to gather staff's opinions as to what information they typically look for. The questionnaire contained three types of questions. The first type was concerned with information relevance. Here, we ask subjects to list the top ten information categories they seek about a fellow staff member. The second type of questions was about current staff information seeking patterns (e.g., how often are you seeking staff information?, what means do you use? and how long does it normally take). The third type of questions was to do with subjects' demographical information. In total, there were 12 questions in the questionnaire. The questionnaire was distributed to 22 randomly selected staff members via email. The subjects were asked to reply within a week's time. One day before the initial due date, a reminder was sent to those subjects who had not yet replied. In the end, ten responses were received.

After receiving the responses, data was collected according to specific question. Regarding to the question of what staff information you are interested to know, over twenty categories of information were registered, including regarding someone's hobbies, and someone's colleagues. Clearly, different people will seek different information. As tailoring is not the focus of this paper, we will not explore those issues further. Instead, we focus on the common subset of staff information categories that emerged from the data set.

Figure 1 shows the top 11 relevant information categories. Over 40% of subjects agree on the top seven categories of staff information, namely, phone number, publications, email address, expertise, (physical) address, experience, and qualification. It is interesting to note that among the top 7 categories of staff information, two of them, expertise and experience, are not explicitly retained by the organisation. The remaining five categories (which we now consider our top 5 categories) come from different data sources in the organisation as shown in Table 1. For example, staff qualification information comes from the human resource system, while publications are stored in the organisation's library database. We realise that, sometimes, the information might already be aggregated in someone's web page; there are, however, wide variations in the information included in staff webpages. Furthermore, a webpage already constitutes a document aggregating information. For our study, we thus explicitly went back to the organisational data sources to find the information, in alignment with the fact that most staff members reported that they use the intranet to find information about someone else.

Staff information categories	**Data sources**
Phone number	Staff look up web page
Publications	Library database
Email	Staff look up web page
Physical address	Intranet
Qualification	Human resource system

Table 1. Data sources for the top five categories of staff information

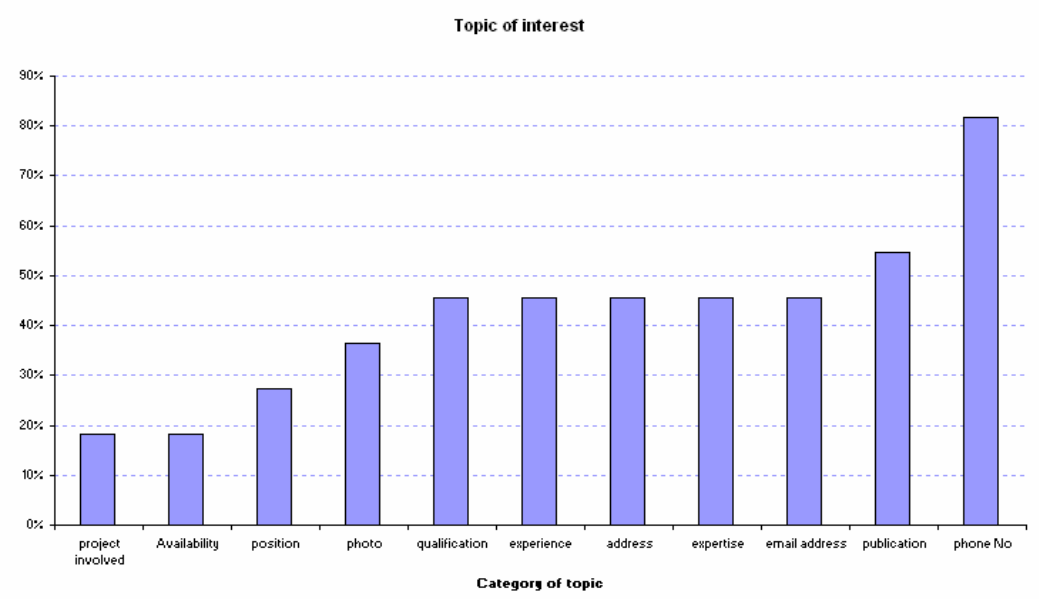

Figure 1. Top 11 staff information categories.

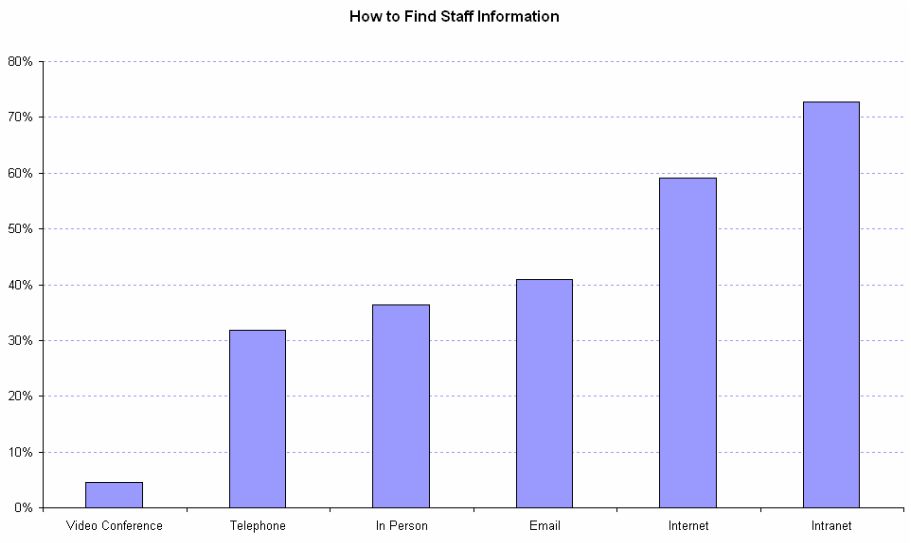

Figure 2. How people go about finding staff information

We briefly summarise some of the patterns that emerged. Overall, a lot of people seek staff information daily or monthly. Most people spend in the order of few minutes for a given time, and they felt that they succeeded in obtaining the information they wanted most of the time. On the issue of how people go about finding staff information, it is probably no surprise to find that the intranet is the preferred means followed by the internet and email (see Figure 2). However, the fact that more people opted for in person than telephone is interesting. We will not elaborate further on these issues as it is not the focus of this paper.

3.Information access efficiency: defining metrics

In this section, we present our approach for quantitatively measuring information access efficiency. In this approach, we differentiate two information accessing modes: inter-item *independent* and inter-item *dependent* accessing mode. Inter-tem independent accessing mode refers to the case when any of one information item is accessed alone. Inter-item dependent access takes into account the fact that items of information to be retrieved are not necessarily independent of each other. For example, they may be obtained using the same application, in which case the time to launch the application should be counted only once. In this case, however, the access point might need to be changed. We will simply refer to the two modes as independent access and dependent access for short.

We then introduce the notion of information access time and speed. The access time of an information item is defined as the time required performing the interaction steps required for accessing it. Under the independent access mode, the access time is the sum of application launch time and other interaction time. Under the dependent access mode, the access time for an information item is the time lapsed between accessing two information items. Under the dependent access mode, the access time for an information item is the time required to access a second item of information once one item has been retrieved.

Access speed is defined as the number of items accessible within a unit of time. Using this definition, average speed can be calculated for the two information accessing modes. The average independent access speed provides a measure of how efficient an application is when accessing any single item, while the average dependent access speed provides a measure of how efficient an application is when any item is accessed after other items. Access speed is then used as a measure of information access efficiency. Our hypothetical information aggregation system is then compared with the current intranet based system using these metrics.

3.1. Establishing steps required to access information

In this part of our study, we used the top five categories of staff information identified in our first experiment as a starting point. For each category of information, we recorded what was needed to be done to find that information: for example, how many tools or applications must be used; how many access points one must go through; how many links need to be selected to get the right information. That is performed for both independent access and dependent access modes. Since most users prefer the intranet, we used the intranet as our choice of information access method.

Information category	load application	key in text	switch access point	click/select hyper link or menu item	press action button or Enter key
Phone number	1	1	0	2	1
Publication	1	1	0	6	1
Email	1	1	0	2	1
Physical address	1	1	1	5	1
Qualification	1	1	0	2	1

Table 2. Number of interaction in the independent access case

3.1.1. Independent access interaction steps

As mentioned earlier, inter-item independent accessing mode refers to the case when any of one information item is accessed alone. Independent access interaction steps for an information item are all the interaction steps required to access the information item in the independent access mode. Taking the phone number as an example, it is accessible with the following steps:

(P1) start the web browser and go to the intranet;
- organisation's homepage appears;

(P2) select the *staff* menu item from the *support* menu ;
- *staff* page appears;

(P3) click the *staff lookup* hyper link from the *Staff* page;
- *staff lookup* page appears;

(P4) key in the last name of the sought after staff member, and

(P5) press *start search* button;
- the phone numbers together with the full names for all staffs with the same given last name are displayed.

After we recorded all the interaction steps required to obtain each information category, we classified them into different types, as each interaction step type requires a different amount of time to execute. We used the following four types: load application, key in text, click/select hyper link/menu item, and press action button.

- **load application:** when an application is launched;
- **key in text:** when text needs to be typed into a text field;
- **switch access point:** when switching to a different web page without hyper link support;
- **click/select hyper link or menu item:** when clicking/selecting a hyper link or menu item;
- **press action button or Enter key:** when pressing an action button or the Enter key.

For each information category, we count how many interaction steps falls into the different interaction types. Table 2 provides a summary of the number of steps required for each interaction step type for the top five information categories if they are accessed independently from each other.

For the phone number example, the number of interactions required are 1, 1, 0, 2, and 1 for interaction type load application, key in text, click/select hyper link/menu item, and press action button, respectively. The 2 interactions required for click/select hyper link/menu item refers to steps (P2) and (P3) above. This is summarised in Table 4 below.

Interaction type	Interaction step
load application	P1
key in text	P4
switch access point	NA
click/select hyper link or menu item	P2, P3
press action button or Enter key	P5

Table 4. Mapping between interaction step and interaction type for accessing phone number in the independent mode

Information category	Load application	key in text	switch access point	click/select hyper link or menu item	press action button or Enter key
Phone number	1	1	0	2	1
Publication	0	1	1	6	1
Email	0	0	0	0	0
Physical address	0	1	2	5	1
Qualification	0	1	1	2	1

Table 3. Number of interactions in the dependent access case

3.1.2. Dependent access interaction steps

Inter-item dependent access takes into account the fact that items of information to be retrieved are not necessarily independent of each other. Dependent access interaction steps for an information item are all the interaction steps required to access the information item, in the dependent access mode. Taking the qualification as an example, it is accessible with the following steps after the phone number has been accessed:

(Q1) Switch to the qualification access point;
- Human resource home page appears;

(Q2) Key in staff ID, and

(Q3) Press the Enter key;
- Staff information page appears;

(Q4) Click on the personal hyper link;
- Sub-personal hyper link appears;

(Q5) Click on qualification hyper link;
- Description of qualification appears.

Similarly, the dependent interaction steps for publication, email and physical address can articulated.

Table 4[1] provides a summary of the number of steps required for each interaction type for the four information items if they are accessed after phone number has been accessed.

3.2. Defining Access Time

In order to measure information access efficiency, we first would like to quantify information access time. To that end, we leverage on research in the field of psychology of human computer interaction. In terms of interaction time, three levels of interaction have been proposed [2]: *psychological moment*, *unprepared response* or *immediate behaviour* [4], and *unit task*. *Psychological moment* is the finest level of interaction at roughly 0.1 seconds. An action and a stimulus event that occurs within 0.1 seconds will seem to exhibit cause and effect relations. *Immediate behaviour* is the intermediate level of interaction at about 1 second. Events that happen in this time frame happen too quickly for the user to respond unless prepared. *Unit task* is the coarse level of interaction ranging from 5 to 30 seconds. This is the typical pace of elementary interaction cycle in interactive systems. An example is the time for a routine interaction with an interactive text editor.

Now, if we can reasonably link our interaction type in last section with the levels of interaction, then we can calculate access time for an information category. Intuitively, the interaction types of **press action button or Enter key** and **click/select hyper link or menu item** corresponds to *psychological moment* and *immediate behaviour*, respectively. For **load application, key in name,** and **switch access point,** they all correspond to the interaction level of *unit task.* However, we think the time designated to *unit task*, namely, 5 to 30 seconds, is too coarse grained for our purpose. So, we introduce three finer levels of distinction to *unit task*: *fine unit task*, at roughly 5 seconds; *intermediate unit task*, at roughly 10 seconds; and *coarse unit task*, at roughly 30 seconds. With this new extension, **key in text** and **switch access point** would correspond to *fine unit task*, and **load application** to *intermediate unit task* (Table 5).

Interaction level		Interaction type	Time interval (sec.)
1	*Psychological moment*	press action button or Enter key	0.1
2	*Immediate behaviour*	click/select hyper link or menu item	1
3	*fine unit task*	key in text, switch access point	5
4	*intermediate unit task*	load application	10
5	*coarse unit task*		30

Table 5. Correspondence between interaction levels and interaction types

Using the information provided in Table 5 above, the access time for an information item may be defined as the sum of the products between required time and number of interaction levels for accessing that item.

$$t(x) = \sum_{i=1}^{5} T_i \, f_i(x) \qquad (1)$$

Where, x -- information item, such as phone number and publications;

$t(x)$ -- access time for information item x;

i -- level of interaction;

T_i -- time required for interaction level i;

$f_i(x)$ -- number of interactions at level i required for accessing item x.

3.2.1. Independent access time

Independent access time is the time required to access an information item in the independent access mode. It can be calculated with formula (1) by replacing $f_i(x)$ with the number of interactions at level i required for accessing item x in the independent mode. For example, the independent access time for

[1] At inter-item dependent access mode, email address will be obtained when accessing phone number. Therefore, no more interaction is required.

information item phone number can be calculated by combining the data provided in Table 5 and first row in Table 2 as follows.

$$t(phone\_no.) = 0.1\times1 + 2\times1 + 5\times1 + 1\times10 = 17.1 \quad (2)$$

Likewise, the independent access time for the rest of the top 5 information items can be calculated. They are shown in Table 6.

Information category	Independent access time (sec)
Phone number	17.1
Publication	21.1
Email	17.1
Physical address	25.1
Qualification	17.1
total	97.5

Table 6. Independent access time for the top five staff information categories

3.2.2. Dependent access time

Dependent access time is the time required to access an information item in the dependent access mode. It can be calculated with formula (1) by replacing $f_i(x)$ with the number of interactions at level i required for accessing item x in the dependent mode. The dependent access time for the top 5 information items can be calculated by using the data provided in Table 5 and Table 4. They are shown in Table 7.

Information category	Dependent access time (sec)
Phone number	17.1
Publication	16.1
Email	0
Physical address	20.1
Qualification	12.1
total	65.4

Table 7. Dependent access time for the top five staff information categories

3.3. Defining Information Access Speed

The information access efficiency of an application/system can be measured by the average information access speed. The higher the speed, the more efficient the system. The information access speed can be defined as the number of information items accessible within a unit of time.

The average information access speed $\bar{v}$ can be calculated with formula (3).

$$\bar{v} = \frac{n}{\sum_{j=1}^{n} t_j} \quad (3)$$

Where, n -- the total number of information items;

t_j -- the access time for the j[th] information item.

3.3.1. Independent access speed

The item independent access speed is the average number of information items accessible within a unit of time under the independent access mode. It can be calculated with formula (3) by replacing t_j with the independent access time for the j[th] information item.

Applying the data shown in Table 6 to formula (3), the average information independent access speed $\overline{v_{in}}$ for the system in which users simply have to use the intranet to gain access to the top five information items is calculated as follows:

$$\overline{v_{in}} = \frac{5}{97.5} \approx 0.051 \text{ (items per second)} \approx 3 \text{ (items per minute)} \quad (4)$$

3.3.2. Dependent access speed

The item dependent access speed is the average number of information items accessible within a unit of time under the dependent access mode. It can be calculated with formula (3) by replacing t_j with the dependent access time for the j[th] information item.

Applying the data shown in Table 7 to formula (3), the average information dependent access speed $\overline{v_{de}}$ for the system in which users simply have to use the intranet to gain access to the top five information items is calculated below:

$$\overline{v_{de}} = \frac{5}{65.4} \approx 0.076 \text{ (items per second)} \approx 4.6 \text{ (items per minute)} \quad (5)$$

3.3.3. Information access efficiency

With the information access speed defined in the preceding section, it is possible to evaluate information aggregation systems in terms of information access efficiency. We can thus consider our hypothetical system, one which would present to users a specific web page constructed on demand, aggregating information from various data sources. Using such a system, users would be able to find in one-go the top five information items required for any staff member, aggregated in a single page. We can now calculate both independent and dependent access speed for this hypothetical system. They are $\overline{v_{in}} \approx 4$ and $\overline{v_{de}} \approx 20$ item/min respectively.

Therefore, we can say that the new system would be about ($\frac{4}{3}$ = 1.3) one point three times more efficient in terms of independent access and about ($\frac{20}{4.6}$ = 4.3) four point three times more efficient in terms of dependent access. Not surprisingly, the gain of aggregation and synthesis is significant only on the dependent case, that is when several items of information need to be found.

Knowing people's information access patterns regarding the number of items normally retrieved at one time would allow us to get a more conclusive evaluation of the time gain of our hypothetical system over simply using the intranet. If the answer is mostly one information item, then there is probably not much point to build the envisaged new system since the new system is only marginally more efficient than the current one (1.3 times efficient). However, if what people look for are all information items at once, then there are more incentives (4.3 times efficient) to develop the new system that provides aggregation.

4. Conclusion

One of the advantages for information synthesis and aggregation is that it is claimed to be easier for end users to access relevant information, especially when information comes from multiple heterogeneous data sources. Although that claim is plausible, it has not been verified by any quantitative studies. It is even unclear how one would quantify information access efficiency to be able to compare various approaches to accessing information. In this paper, we define a way to measure information access efficiency which is grounded on information access speed. Information access speed for any information provision system can be determined by following a set of steps.

- identify interaction steps for accessing information items;
- map interaction steps to interaction levels (time);
- calculate information item access time with the access time formula;
- calculate information access speed with the access speed formula.

The average information access speed is then used as criteria for assessing information access efficiency of specific applications in the context of presenting end users with aggregated staff information. This specific study provided us with a simple case where information aggregation and synthesis improves information access efficiency.

To conclude, the significance of our work is three fold. First, to our knowledge, it represents a first attempt to study information access efficiency empirically. It is a step forward in the value proposition of information synthesis and aggregation from qualitative to quantitative measures. Second, our study provided a simple example where it is more efficient to access staff information items when they are aggregated into a single page than accessing them from multiple access points. Third, our study provided an easy and practical approach for gauging whether developing a particular information aggregation application is worthwhile.

Acknowledgements

We wish to thank Ross Wilkinson for originating the ideas of the information access efficiency experiment. We also thank George Ferizis and the anonymous reviewers of the paper for their useful comments. Finally, we thank the people who participated in our survey.

References

[1] Campbell, M.K., DeVellis, B.M., Strecher, V.J. Ammerman, A.S., DeVellis, R.F. and Sandler, R.S. (1994): Improving dietary behavior: The effectiveness of tailored messages in primary care settings. American Journal of Public Health, 84:783–787.

[2] Card, S.K.; Moran, T.P.; and Newell, A. (1983): The psychology of human-computer interaction. Hillsdale, NJ: Lawrence Erlbaum Associates.

[3] Koedinger, K.R.; Anderson, J.R.; Hadley, W.H. & Mark, M.A. (1997): Intelligent tutoring goes to school in the big city. International Journal of Artificial Intelligence in Education, 8, 30-43.

[4] Newell, A. (1990): Unified theories of cognition. Cambridge, MA: Harvard University Press.

[5] Paris, C., Wu, M., Vander Linden, K., Post, M. and Lu, S. (2004). Myriad: An Architecture for Contextualized Information Retrieval and Delivery, in AH2004: International Conference on Adaptive Hypermedia and Adaptive Web-based Systems. August 23-26, The Netherlands. pp. 205-214.