

# Multi-resolution information transmission in mobile environments\*

Hong Va Leong<sup>a</sup> and Antonio Si<sup>b</sup>

<sup>a</sup>*Department of Computing, Hong Kong Polytechnic University, Hong Kong*

<sup>b</sup>*Oracle Corporation, CA, USA*

**Abstract.** Mobile environments are characterized by low communication bandwidth and frequent disconnection. Conventional information retrieval and visualization mechanisms thus pose a serious challenge to mobile clients. There is a need for these clients to quickly perceive an overall picture of the information available to them, so as to enable them to discontinue the transmission of information units that are unlikely useful to them. We had proposed a *multi-resolution transmission mechanism* for web documents. In particular, various organizational units of a document are transmitted to a mobile client in an order according to their information content, thereby allowing the client to terminate the transmission of a useless document at an earlier moment. In this paper, we generalize the *multi-resolution transmission model* for a document, and then extend that model into the multi-resolution transmission framework to cater for not only units within a document, but also for a collection of documents. We refer to the multi-resolution transmission mechanism for a particular document as *intra-document multi-resolution transmission mechanism* and the extension to a document cluster as *inter-document multi-resolution transmission mechanism*. With the integrated multi-resolution transmission framework, a mobile client can examine the important portions of the document cluster for an early grasp of the information therein, with the most important contents for each of those documents more readily available as well.

## 1. Introduction

Recent advancement in wireless communication technologies has revolutionized information sharing by removing the inherent constraint of the wire. Clients carrying a portable device can easily gain access to information stored over the Internet. Information access on the road for mobile clients holding mobile devices is becoming more and more important [17]. Coupled with the rapid miniaturization of computing devices, we have witnessed the proliferation of handheld devices that are becoming smaller in size and more powerful in ability. These handheld devices, ranging from cellular phones to PDAs, to even smart cards, are being deployed in an increasingly aggressive manner. For instance, PDAs grow from merely providing calendar, phone book and scheduler to active information retriever, with which the concept of mobile web has been embraced. Instead of downloading electronic books from wired network and synchronized onto the devices, these resources are increasingly being accessed online and on-demand. Cellular phones extend their usability from voice communication devices to support SMS, MMS and now become Java-enabled with their own OS for task scheduling and resource management. These two types of devices are converging to play the role of classic mobile clients, which normally refer to those laptop computers in the past. When “wireless networks” meet “handheld devices”, the vision of ubiquitous computing [19] is materialized.

---

\*This research is supported in part by the Research Grant Council under grant number PolyU 5084/01E.

We focus on a mobile environment in which mobile clients navigate and visualize inter-related web documents, stored at or routed through a stationary server (or proxy gateway), with common browsers. In particular, the information being accessed is assumed to be closely related to the XML format. Communication between mobile clients and the stationary server is often via low bandwidth wireless channels [2,10]. Furthermore, mobile clients are constrained by limited battery life. Caching of data items from the server in a client's local storage has been investigated [5] to reduce wireless traffic and the dependency on the wireless channels. With the same token, the traffic generated for mobile document browsing should consume as little bandwidth as possible. In this aspect, conventional approaches to web navigation and information visualization suffer from serious limitations.

Conventional approaches to web navigation usually involve the searching of web documents via some search engines, through search structures like smart index [8], Lycos [16] or WebCrawler [18], which typically search exhaustively. After the initial set of web documents is returned, the mobile client will likely explore manually each document for relevance. A probably better approach is to establish a user profile that captures the interests of individual users. The profile is used to filter out irrelevant information identified by a search engine [1,4]. Conventional searching techniques are often employed to identify information in response to a query, with additional filtering techniques to discard irrelevant information according to the personalized profile of the user. Relevance feedback plays an important role in modifying the user profile to reflect the changes in user interest. Rather than providing a user with selected documents, the WebWatcher system [3] assists a user in his/her browsing behavior. It interactively offers advice about which subsequent hyperlink(s) would likely contain the most relevant information. Furthermore, the system refines its knowledge on user interest by keeping track of whether its advice to the user is followed. Concept hierarchies have been employed to improve the organizational power and browsing efficiency for a resource-limited mobile client with small display area [6].

A document returned from web access or searching may simply be a single web page, or a collection of hierarchically linked related pages, composing a larger document. To reduce bandwidth consumption for transmitting a single document to a mobile client with limited display area, formatting information that cannot be visualized due to display limitation can be removed [9]. Similarly, complicated formatting can be simplified. Segments that carry little information can also be filtered or condensed, and the structure of document is adjusted accordingly. A collection of documents returned can often be viewed as a larger logical document. A user exploring those related pages can only traverse the pages following the link he/she thinks to lead to the most relevant information. This is particularly problematic with a mobile client with limited resource. The problem of locating a relevant web page without making many unnecessary traversals is partly alleviated in carefully designed web sites, where there are road maps to different sections of the web page collections, as in Yahoo. Owing to the dynamic and come-and-go nature of the web pages, the accuracy and freshness of these road maps may be greeted with doubt. Related documents returned together from a search engine, or collected together via relatively strong hyperlinks can be consolidated into a more coherent piece of information, by means of some clustering algorithm [11] or inter-page linking [13]. The cluster can be captured in a meta-data structure, in the form of XML documents, as a kind of road map.

We observe quite often that most documents identified by a search engine and sometimes those contained in a document cluster are irrelevant to a user. This not only wastes precious bandwidth for transferring the documents, but also consumes the limited energy of a mobile client unnecessarily. This problem becomes more serious when the size of a logical web document is getting bigger; many of those documents represent technical documents or specifications. A mechanism to allow early termination of transmission of irrelevant documents is needed. In [21], document summarization is performed on

individual documents so as to return a smaller amount of overview information before the relatively large document is actually retrieved. Since the summary is not reusable when the full document is transmitted, extra bandwidth is consumed when the document is really useful. On the other hand, to improve the usability of the document cluster road map for a collection of documents, high-level information about potentially important documents should be transferred to the client. This allows the client to gain an overview of those documents before he/she actually clicks on the link to request for their transmission. This is achieved again at the expense of more bandwidth consumption. However, this tradeoff is worthwhile, since without the high-level information for important documents in the cluster, there is a high chance that the client will initiate a request to download the document for browsing anyway. Actually, the savings gained from early termination of useless document transmission can more than offset the cost to transfer the high-level details of documents, not to mention the gain achieved by reducing the need to actually initiate the transmission of useless documents, as inferred from the high-level information. This calls for the provision of a high quality document overview. In this paper, we propose a *multi-resolution transmission framework*, which embraces these concepts, through consideration of two types of multi-resolution document transmission mechanisms.

We investigated into the paradigm of multi-resolution transmission in [22]. It is concerned with the transmission of a web document in an order that maximizes the expected value to a mobile client. As a result, we refer to this scheme as the *intra-document multi-resolution transmission mechanism*, since it is targeted at a logical web document, rather than a collection of related documents. According to the structure of a document, it divides a document into multiple organizational units (also known as *information units*), according to various Levels Of Detail (LOD). The *information content* for each unit indicates the amount of information in the unit, so that units of higher information content are transmitted earlier. A document can then be transmitted and browsed at a coarser resolution, with the details to be filled in progressively. This has a similar flavor as JPEG compression for images, where significant coefficients are transmitted earlier and insignificant ones are transmitted later, or even dropped without much impact to the image quality. The idea of multi-resolution LOD has also been applied recently on the visualization of spatial data [7]. With multi-resolution transmission, a user could decide early to discontinue with the transmission of an irrelevant document. In this paper, we first generalize the mechanism for information unit transmission within a document. We then extend the notion of information content into different variants beyond *relative information content* [22], *query-based information content* and *modified query-based information content* [12]. The various notions of information content can be used to model the importance of web documents contained in a document cluster graph [14]. We refer to the cluster level realization of the multi-resolution transmission scheme the *inter-document multi-resolution transmission mechanism*. This extended mechanism involves the determination of important documents for which the more important information units should be transmitted, based on the document cluster graph and new notions of information content.

The remainder of this paper is organized as follows. In Section 2, we describe the architecture of the multi-resolution transmission framework. The two multi-resolution transmission mechanisms are explained and discussed in Section 3, with examples. Finally, we offer brief concluding remarks, with an outline of our future research directions in Section 4.

## 2. Multi-resolution transmission framework

Our multi-resolution transmission framework is based on the idea of *Level Of Detail* or *LOD*. For each document, the LOD ranges from document, section, subsection, and so on, down to paragraph. Above

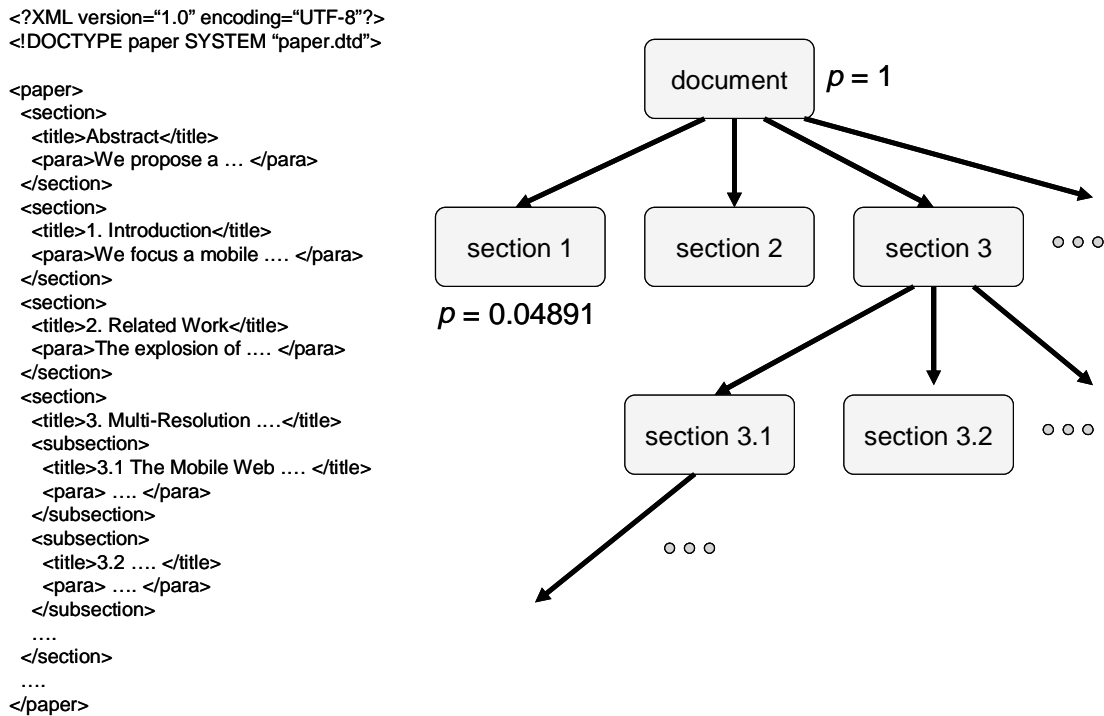


Fig. 1. Sample document and document tree.

the document, we have the document cluster (cluster) and cluster of document clusters (supercluster). As a result, a total order can be established: paragraph  $\triangleright \dots \triangleright$  subsection  $\triangleright$  section  $\triangleright$  document  $\triangleright$  cluster  $\triangleright$  supercluster, where the document level represents the partitioning point between intra-document transmission and inter-document transmission. This is because a document is assumed to be self-contained and any granularity above document level serves mostly for the purpose of organizational convenience, whereas the granularity below document level represents finer fragments of the document as a window for visualization.

For intra-document transmission, each web document can be represented by a semantic structure, called the *structural characteristic (SC)*. It is a tree-like structure modeling the structural organization of a document. Figure 1 illustrates a sample document and its corresponding document tree, defining the SC. Each node  $n_i$  is associated with an information content  $p_i$ , which indicates the amount of information content captured within the organizational unit that it models. Here, the term *information unit* refers to the organizational unit of interest to a particular scenario. It may be the paragraph, the section, or the subsection. There are also different notions of information content, focusing on different aspects of the amount of information contained in an information unit. They may measure the absolute amount of information or the relative amount of information with respect to information unit size or query or both.

For inter-document transmission, the collection of documents currently of potential interest, as returned by a search engine or a document organizer [11], is called a *document cluster*, which can be represented by another semantic structure, called the *document cluster graph (DCG)*. Unlike the SC, the DCG is a graph-like structure, modeling the relationship between a logical collection of documents. Figure 2 illustrates a sample document cluster graph. Each node  $\nu_i$  is associated with a mass  $m_i$ , which represents the amount of information contained in the document  $D_i$  that  $\nu_i$  represents, with respect to the querying

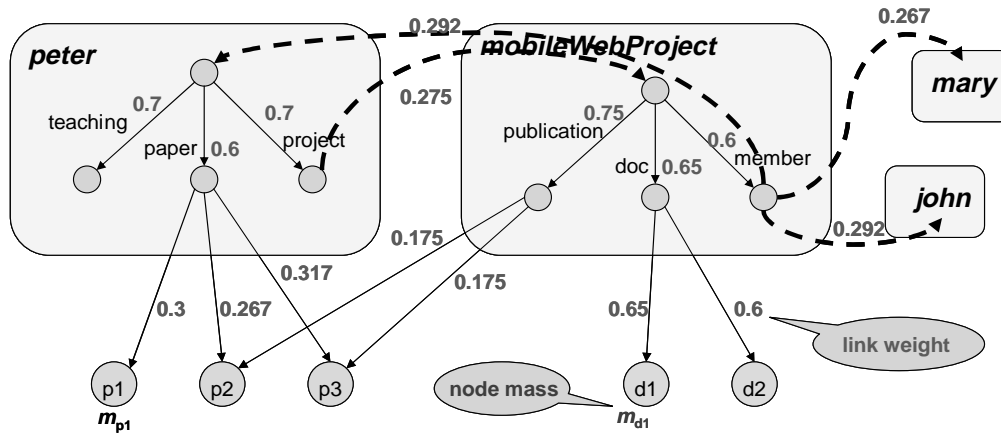


Fig. 2. Sample document cluster graph.

or categorical criteria. The weights of the links represent the affinity or attraction between the nodes. The higher the weight  $w_{ij}$  of a link  $e_{ij}$ , the more chance will node  $\nu_j$  be accessed when node  $\nu_i$  is currently being viewed. As with intra-document transmission, the term *information unit* refers to a node or a cluster for transmission. Transmission order of information units can also be determined based on the notion of information content used in defining the mass of nodes and the weight of links, both factors being influential on the attraction forces between different pairs of connected nodes.

The high-level system architecture is illustrated in Fig. 3. A mobile client initiates a request for a web document or a collection of documents via a web browser or a visualization program, called the viewer. The viewer interacts with the base station which functions as the proxy gateway (proxy). The request is then routed to the appropriate web server or document server for the appropriate documents. Meanwhile, the DCG could be constructed dynamically, retrieved from the target site or from the local cache [11]. After the document or collection of documents (or their summarization or distillation) is received, the proxy will arrange for the transmission of various information units of the requested document, as well as for the transmission of the DCG and the prioritization of relevant information within the DCG. There are four major components at the proxy: request handler, document transmitter, structural characteristic generator and document cluster graph generator. The structural characteristic generator and document cluster graph generator generate the SC and DCG respectively. The transmission decision is generated and managed by the document transmitter, which either retrieves the SC and DCG from the local cache, or requests the appropriate generators to produce the semantic structures. The request handler will intercept client request, forwarding it to the web or document server if the results are not already cached, and storing the relevant information in the cache for future use by the two generators.

There are two major components with each mobile client: sequence manager and rendering manager. The sequence manager is responsible for receiving an information unit of the requested document from the proxy. It will reconstruct the partial document tree for the currently viewing document, through a hierarchical naming system, following the technique of the XNS (XStream Naming Space) in XStream [20], by attaching the received information unit or collection of units to the appropriate position. Similarly, for information pertaining to nodes in DCG, the information unit will be associated with the corresponding node. The change in the document tree of DCG will be passed on to the rendering manager, which renders the unit in a web browser or the viewer at the appropriate position.

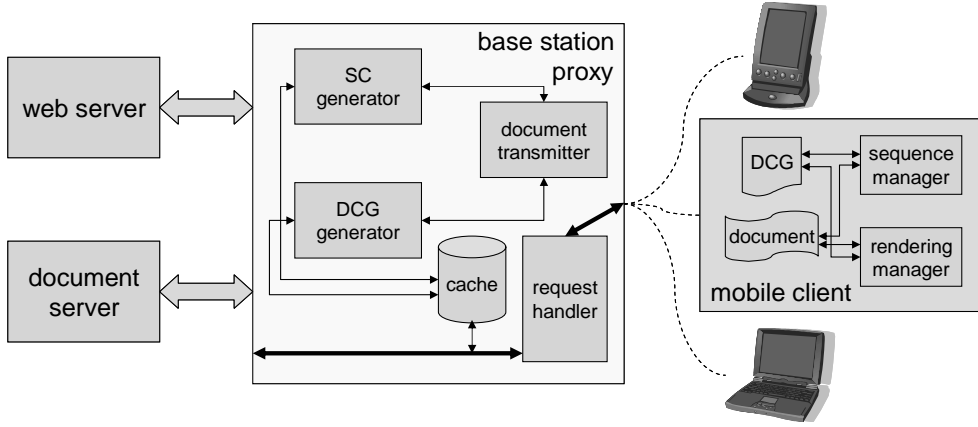


Fig. 3. Architecture of multi-resolution transmission framework.

### 3. Multi-resolution transmission mechanisms

In this section, we will first define the different notions of information content, on which the multi-resolution transmission mechanisms are developed. The two different levels of multi-resolution document transmission are then described in details. An example is also presented to illustrate the mechanisms.

#### 3.1. Information content

The information content  $p_i$  of an information unit  $n_i$  is measured by the amount of information contained in the unit. The amount of information is reflected by the distribution of the keywords in the unit. Keywords which occur more frequently than others in a unit should carry more information of the document, but each occurrence of those high frequency keywords will be less significant, i.e., carrying less information. Each keyword  $a$  is associated with a weight  $\omega_a$ . Non-keywords are considered to possess a weight of zero. Each particular occurrence of  $a$  may be in different context and thus warrant different contribution. For instance, an occurrence inside the abstract or inside a section heading ( $\langle \text{section} \rangle$ ) should be considered more important. We associate a *context adjustment score*  $C_T$  with each occurrence of  $a$  in context  $T$ , defaulted to 1. For example, the context adjustment score for those keywords within  $\langle \text{section} \rangle$  (i.e., section title) may be  $C_{\text{section}} = 3$ , thus allowing the more important constructs to be selected earlier than the less important ones. In other words, we treat the keyword  $a$  staying within an important context  $T$  as to have occurred a multiple of  $C_T$  times.

For notational convenience, we denote the number of occurrences of a keyword  $a$ , after contextual adjustment in a document  $D$ , by  $|a_D|$  and the number of occurrences of  $a$  after contextual adjustment in an information unit  $n_i$ , by  $|a_{n_i}|$ . The occurrence vector  $V_D$  of the set of keywords in  $D$ ,  $A_D = \{a | a \text{ is a keyword in } D\}$ , can be represented as  $V_D = \{|a_D| | a \in A_D\} = \{\nu_1, \nu_2, \dots, \nu_{|A_D|}\}$ . We use a logarithmic function to define the weight  $\omega_a$  of each occurrence of keyword  $a$  in  $D$ . This function could model the decay in importance of  $a$  with respect to its increase in frequency in  $D$ . It is defined as  $\omega_a = 1 - \log_2(|a_D|/||V_D||)$ , where  $||V_D||$  is the norm of the occurrence vector  $V_D$ . We adopt the infinity norm  $||V_D||_\infty = \max(\nu_i)$  here. Using infinity norm, we observe that in a special case where all keywords occur only once, the norm of the occurrence vector will be 1. We can now define the notion of *Plain Information Content (PIC)* more formally.

**Definition 1.** The *plain information content*,  $p_i$ , of information unit  $n_i$  is  $p_i = \frac{\sum_{\forall \text{keyword } a \in n_i} |a_{n_i}| \omega_a}{\sum_{\forall \text{keyword } d \in D} |d_D| \omega_d}$ .

PIC is defined to indicate the amount of information contained in an information unit, without concerning on the size of the unit. In a mobile environment, the tradeoff between amount of information content and the time and amount of bandwidth consumed needs to be considered. We can thus define the notion of *Relative Information Content (RIC)* to take into account this content and size tradeoff, with a goal of accumulating more information within the shortest possible duration at a client. We have shown in [22] that transmitting information units in descending order of RIC will maximize the expected amount of information received by clients. To define RIC, let the size of information unit  $n_i$  be denoted as  $l_i$ .

**Definition 2.** The *relative information content*,  $\gamma_i$ , of information unit  $n_i$  is the ratio of its plain information content to its length, i.e.,  $\gamma_i = p_i/l_i$ .

The notions of PIC and RIC are based on a static analysis of a document. Documents transmitted to a mobile client, to be browsed by a user, are often generated by a searching process via some search engines. The degree of relevance of a document, and thus its information units, is usually affected by its relevance to the query initiated. We extend the definition of PIC to *Query-based Information Content (QIC)* to take into account of an issued query. Note that while the PIC of an information unit is static, its corresponding QIC is dynamic, changing according to an initiated query due to the presence of extra information in the keyword-based query.

We denote a query  $Q$  by a vector, analogous to that of a document. The query  $Q$  contains a set of keywords, which we call the *querying words*,  $A_Q = \{a | a \text{ is a keyword in } Q\}$ . This forms an occurrence vector  $V_Q$  for  $Q$ . We could also take the weight of each querying word into account, in parallel to documents. The weight of a querying word  $a$  is  $\omega_a^Q = 1 - \log_2(|a_Q|/|\nu_Q|)$  if  $|a_Q| \neq 0$  and it is zero otherwise. The QIC  $q_i^Q$  for a document  $D$  and the query  $Q$  is the combined weighted sum of the keywords in the unit, normalized with respect to  $D$  and  $Q$ . Formally, we have the following definition for QIC.

**Definition 3.** The *query-based information content*,  $q_i^Q$ , of information unit  $n_i$  with respect to query  $Q$  is  $q_i^Q = \frac{\sum_{\forall \text{keyword } a \in n_i \cap Q} |a_{n_i}| \omega_a \omega_a^Q}{\sum_{\forall \text{keyword } d \in D \cap Q} |d_D| \omega_d \omega_d^Q}$ .

One of the disadvantages of QIC is that the information content of an information unit may become zero, due to the absence of a querying word. This can be avoided through a modification of the definition of QIC with a scaling factor  $\rho$  to convert the product between document keyword and querying word into a sum. This is called the *Modified Query-based Information Content (MQIC)*  $\tilde{q}_i^Q$ , which can be defined formally.

**Definition 4.** The *modified query-based information content*,  $q_i^Q$ , of information unit  $n_i$  with respect to query  $Q$  is  $\tilde{q}_i^Q = \frac{\sum_{\forall \text{keyword } a \in n_i} |a_{n_i}| (\omega_a + \rho \omega_a^Q)}{\sum_{\forall \text{keyword } d \in D} |d_D| (\omega_d + \rho \omega_d^Q)}$ , where  $\rho = \frac{\sum_{\forall \text{keyword } a \in D} |a_D|}{\sum_{\forall \text{keyword } a \in Q} |a_Q|}$ .

Once again, we can define the notions of *Relative Query-based Information Content (RQIC)* and *Relative Modified Query-based Information Content (RMQIC)* to take into account the balancing need of transmitting a higher information containing unit in the shortest moment, by taking on the ratio of the respective information content definition with the size of that unit.

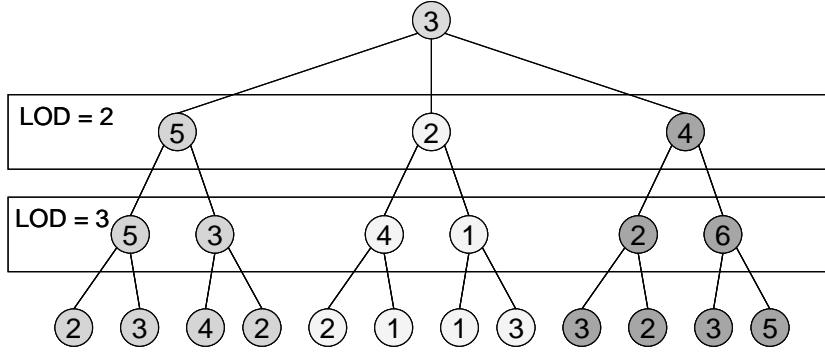


Fig. 4. A document for intra-document transmission.

**Definition 5.** The *relative query-based information content*,  $\gamma_i^Q$ , of information unit  $n_i$  with size  $l_i$  with respect to query  $Q$  is  $\gamma_i^Q = q_i^Q / l_i$ .

**Definition 6.** The *relative modified query-based information content*,  $\tilde{\gamma}_i^Q$ , of information unit  $n_i$  with size  $l_i$  with respect to query  $Q$  is  $\tilde{\gamma}_i^Q = \tilde{q}_i^Q / l_i$ .

### 3.2. Intra-document transmission

Intra-document multi-resolution transmission is focused on the mechanism to transmit a logical document to the mobile client. The client can choose the appropriate LOD for which the document is to be transmitted. The nodes in the SC of the document at that LOD, namely, a traversal of nodes at that particular LOD, are collected. The chosen form of information content (e.g., PIC, RIC, or RMQIC) is then computed for each of these nodes. The corresponding information units are ordered according to the selected information content for transmission. Transmitting units with higher information content earlier, the client can perceive a higher level of content early on.

There are different mechanisms to transmit sub-units contained within an information unit to the client. In our previous work, we elect to transmit sub-units for a chosen information unit in a *sequential manner* [22]. In this paper, we generalize the transmission order of those sub-units so that they could follow the order according to their information content and this can occur in a recursive manner. The former conventional approach is referred to as *flat LOD transmission* and the latter generalized approach is referred to as *hierarchical LOD transmission*. The advantage of flat LOD is the simplicity and intuitive client perception. This is because when a client selects a high LOD, he/she may have expected that the transmission granularity is at that particular LOD, with details of information units at that LOD transmitted orderly and progressively. Flat transmission ensures that the perceived granularity and its orderliness is not violated, since all information units at that granule are received before others and in the same order.

In hierarchical LOD transmission, the transmission granularity is dynamic, so as to maximize the effectiveness of multi-resolution transmission in terms of user perceived amount of information content. Here, the sub-units for an information unit that is selected to be transmitted at the chosen LOD, called the *upper LOD limit*  $\mu$ , are no longer transmitted sequentially. Instead, those sub-units are ordered according to their appropriate variant of information content. As a result, the expected information content observed by a client can be maximized under the constraint that eventually, all sub-units of the specific information



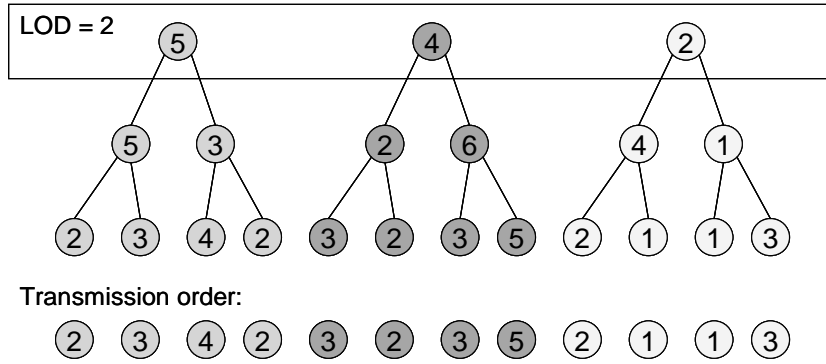


Fig. 5. Flat LOD transmission order at section level.

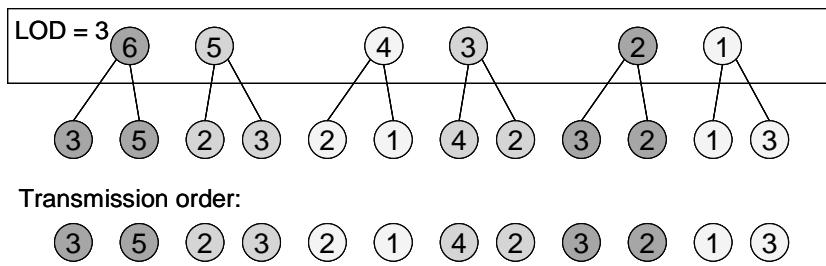


Fig. 6. Flat LOD transmission order at subsection level.

unit are transmitted before others. In other words, all nodes of the sub-tree rooted at the selected unit are transmitted in a consecutive stream, before the nodes under other untransmitted sub-trees. Thus, the granularity of the unit can still be preserved, but the orderliness of the unit is violated, since sub-units under the chosen unit are transmitted in an intermixed manner. To further generalize the hierarchical LOD transmission model, we will allow the client prescription of another LOD as a lower bound. This new lower bound LOD, called the *lower LOD limit*  $\lambda$ , will define a limit on the recursive ordering of sub-units for transmission, beyond which all descendant units are transmitted sequentially. Thus, this defines a recursion basis as a restriction on the recursive hierarchical LOD transmission scheme.

In fact, both flat LOD transmission and hierarchical LOD transmission can be unified into one single model with the two LOD parameters: upper LOD limit  $\mu$ , and lower LOD limit  $\lambda$ . Our conventional multi-resolution transmission mechanism can be considered a special case for which  $\mu = \lambda$ . If one chooses  $\mu = \lambda$  at document level, the whole document will be transmitted sequentially, equivalent to conventional TCP or HTTP document transmission.

To illustrate the intra-document multi-resolution transmission scheme with the two LOD limits,  $\mu$  and  $\lambda$ , we consider the document in Fig. 4, where the values in the nodes represent their information content. The three branches or three sections of the document bear different shading for exposition purpose. With flat LOD transmission, the limits  $\mu$  and  $\lambda$  are equal. The transmission order with  $\mu = \lambda = 2$  and  $\mu = \lambda = 3$  are depicted in Figs 5 and 6 respectively. Here the third branch is transmitted before the second at section LOD in Fig. 5, due to its higher information content. Similarly, branches at subsection LOD are rearranged in Fig. 6. Sub-units within the same branch are transmitted sequentially in both cases.

With hierarchical LOD transmission, the default value of  $\lambda$  is at paragraph level. As exemplified in Fig. 7, with  $\mu = 2$ , the sections are ordered similar to that in Fig. 5. Here the white nodes represent those

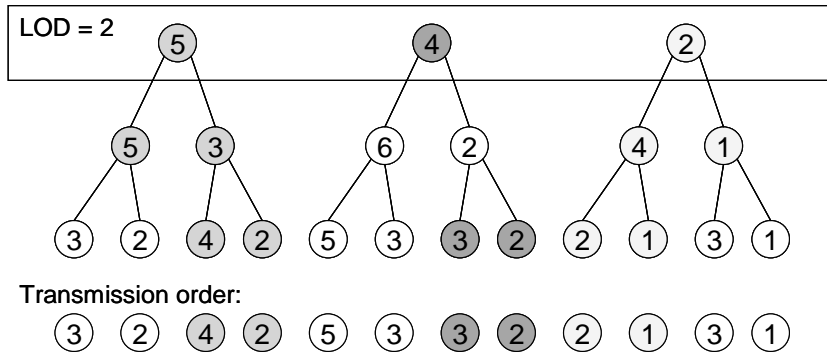


Fig. 7. Hierarchical LOD transmission order down to paragraph level.

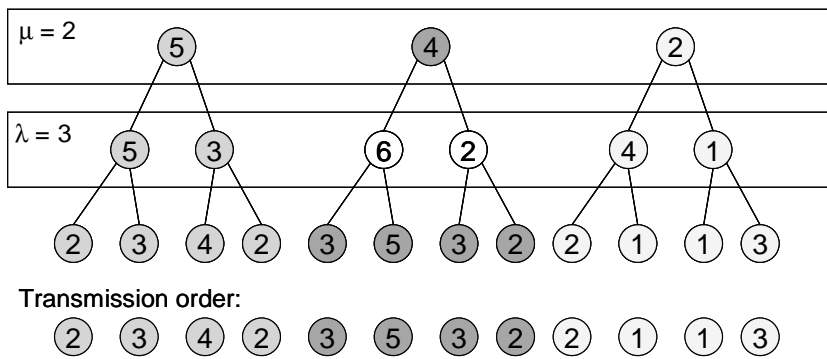


Fig. 8. Hierarchical LOD transmission order down to subsection level.

sub-units experiencing re-ordering. Furthermore, sub-units for each branch are re-ordered recursively according to the information content of the corresponding nodes. Finally, when  $\lambda = 3$ , re-ordering of sub-units terminates at subsection level with the remaining sub-units transmitted sequentially, as depicted in Fig. 8.

### 3.3. Inter-document transmission

Inter-document multi-resolution transmission is focused on the mechanism to transmit general information about documents contained in a document cluster to the mobile client. The client can then gain an overview of the collection of documents in the cluster. Based on the information content contained in constituent documents in the document cluster or supercluster, the mass for each constituent document can be computed. Based on the access needs of the documents, weights of the links between the documents can be computed. The nodes in the DCG of a cluster are collected and the overview information are transmitted to the client in the order of importance, in anticipation of the corresponding document being accessed. Let the DCG be represented by  $G = \langle V, E \rangle$ ,  $V = \{\nu_1, \nu_2, \dots, \nu_{|V|}\}$  and  $E = \{e_1, e_2, \dots, e_{|E|}\}$ . An edge  $e_k \in E$  represents a link connecting nodes  $\nu_i$  and  $\nu_j$ , also denoted  $e_{ij}$ . Nodes with higher information content are likely more useful than those with lower information content and nodes with stronger links with the currently viewing node (without being terminated early) are also of higher potential. Inter-document multi-resolution transmission mechanism needs to prioritize the nodes and arrange for the top-level information units to be transmitted as a form of prefetching.

For instance, the top-level of a technical report would well be Section 0.0.0, following our convention of naming information units [22]. For most cases, it would be the title page or the abstract, and is sufficiently indicative of the contents that the whole document contains.

In conventional document cluster graph browsing [14], only the title of each document is transmitted to the client. High-level document summaries are under consideration for transmission only when they are of a small size. When the client selects a document, the low-level information units are transmitted with the intra-document multi-resolution transmission mechanism. With inter-document transmission, the high-level summary or the high-level information units of each document could be transmitted to the client in form of prefetching, especially when a client is busy viewing an existing information unit. This prefetching pays off if the chance that the high-level information is needed is comparatively high or it is potentially not useful to transmit remaining information units of the currently viewing document.

In this paper, we assume for simplicity that the streams generated by the multi-resolution transmission of document cluster graphs and documents, i.e., inter-document and intra-document transmission, will be multiplexed by the proxy. In other words, the two mechanisms can be executed on two logical channels, in an independent manner. This arrangement simplifies the design of the mechanisms. We would leave the research issue of the analysis of the breakeven point of prefetching against transmitting information units of existing document as future work.

There are two major parameters in a DCG, namely, the weight of links and the mass of nodes. The weight of a link represents the affinity or attraction between the two nodes at its endpoints and the mass of a node reflects the amount of information it contains. The mass  $m_i$  of a node  $\nu_i$  can simply be the plain information content or other notions of information content for document  $D_i$  that it represents. The higher the mass, the more important or more informative the document is. These two parameters are explained in further details in subsequent sections.

### 3.3.1. Link weight in document cluster graph

The weight of a link is usually dependent on two factors: access frequency of the two nodes and the semantic distance between the nodes [11]. Semantic distance measures the closeness in content or access affinity or probability of being co-referenced. For documents returned by a search engine, the semantic distance may well be the simple similarity score returned by the vector space model. For a document cluster, the semantic distance may be represented by the network distance that indicates the geographical relationship between nodes, as well as its variant. For attraction cluster formation [11], we define the weight  $w_{ij}$  of link  $e_{ij}$  between nodes  $\nu_i$  and  $\nu_j$  as  $w_{ij} = \alpha Pr(\nu_j|\nu_i) + (1 - \alpha)D_{ij}$ , with an adjustable system parameter  $\alpha \in [0,1]$ . In the definition, the conditional access probability of the documents represented by the nodes can be approximated by the access frequency. The semantic distance  $D_{ij}$  is computed as  $1/d^\theta$ , where  $d$  is the number of steps (using Unix “cd .” and “cd subdir” commands) to traverse from the directory containing  $\nu_i$  to the directory containing  $\nu_j$  and  $\theta$  is another adjustable system parameter that governs the strength of relationships. It is common to choose  $\theta = 1$  for the inversely proportional property. The distance  $d$  will be one when both nodes reside in the same directory, and it is a large value when the nodes lie across two administrative domains. The number of steps going from one domain to another is often a reasonably large value of, say, 5 to 10.

### 3.3.2. Document mass in document cluster graph

Defining the mass simply as the plain information content does suffer from several limitations. First, the definition of PIC implies that the *total* information content of a document is always one. All documents then become equivalent. Second, the length or size of the document is not taken into consideration. Third,

the quality of the title page or abstract as the top-level information unit is not reflected by different notions of information content. A good abstract is more useful than a badly-written one and a long abstract can carry more information than a short one, but is also more expensive to prefetch. Furthermore, it may not be always the case that the top-level information unit represents the document in the best way. The abstract may lie in section 1 instead of section 0 and a poorly-written abstract may be outshot by the conclusion or other sections. Finally, each document may be associated with a distilled description which may be generated by some summarization algorithm [15]. Such a description is often more useful than just the top-level information unit, but it is not part of the document, and does not enter into the computation of information content. This is particularly important for documents that do not resemble the structure of a technical report, i.e., the top-level information unit is neither the title page nor the abstract.

To address the first issue, it is more appropriate to adopt other variants of “total” information content to represent the mass of the document, with the intuition that a larger document should contain more information, despite being more lengthy. The second issue can be addressed with RIC and its query-based variants, by taking into account information content and size tradeoff. Given the limitation of information content being normalized to a value of 1 for any document, RIC as defined upon PIC can only prioritize the units within the same document, but not across the documents. A variant of RIC with respect to total information content would be more appropriate. This notion is defined in Section 3.3.3, called *information density*,  $\Gamma$ .

As with the third issue, the utility metric for prefetching an information unit  $n_j$  of document  $D$  should be tied to the total information content of the top-level unit  $n_i$  or the chosen unit  $n_j$ . The quality of an information unit  $n_i$ , for example, the abstract or the summary, is reflected by the information density  $\Gamma_i$  of that unit. To be more precise, we would like to measure the quality of the unit with respect to other units in the document. We thus define the *quality*  $\psi_i$  of unit  $n_i$  within  $D$  to be the effectiveness of the unit in representing the overall information in the document. It is the ratio of  $\frac{P_i}{P_D}$  to  $\frac{l_i}{l_D}$ , which can be simplified into  $\psi_i = \frac{\Gamma_i}{\Gamma_D}$ . Prefetching for each document should be based on this information density metric or quality metric.

Finally, the quality of a distilled section should normally be higher than that of an ordinary section and a manually written one is even better. We thus place a higher priority to summary returned by external algorithms or defined explicitly by users, under the belief that it is an effective distillation of the document. There is a need to put a bias towards such kind of information units. Recall that the best information unit can be determined as the one with the highest quality or highest information density. To enhance the competitiveness of a summary unit, we could apply a biasing factor to refine the metric for each document, before a global ordering for transmission. The modified metric for document  $D$  is thus  $\psi^* = B \times \psi$  or  $\Gamma^* = B \times \Gamma$ , where  $B \geq 1$  is a boost factor. We would choose  $B$  to be a moderate value if the unit is generated by a human and a smaller value if it is machine-generated.

So far, we have not utilized the weight of links in computing the transmission order. It is also conceivable that clients would visit the nodes with a probability proportional to the weight of the links. That is the basis on which the document cluster was generated. Once the document cluster is generated in the form of a DCG, a client can already perceive the DCG as a whole. The weights of the links should only play a lighter role on the degree of importance of the top-level abstraction or summarization to be prefetched. As such, the weights can be accumulated at a discounted rate to the mass or the quality metric for transmission order determination.

### 3.3.3. Extended notions of information content

The notion of total information content is useful to place a global ordering on information units across documents. This is realized by the notion of *Absolute Information Content (AIC)*, which measures the actual amount of information contained in a document, without being normalized within the document. The weight of a keyword  $a$  is thus homogeneous across documents. It may be defined according to the one used for a document as  $\omega_a^* = 1 - \log_2(|a_{D^*}|/||V||)$  with  $||V_{D^*}||$  being the norm of the occurrence vector  $V_{D^*}$ , by considering the whole cluster as one large logical document, i.e.,  $D^* = \cup_i D_i$ .

**Definition 7.** The *absolute information content*,  $P_i$ , of information unit  $n_i$  is  $P_i = \sum_{\forall \text{keyword } a \in n_i} |a_{n_i}| \omega_a^*$ .

This simple definition is highly computationally inefficient. When the set of documents changes, the weights need to be recomputed. It would be more efficient if one adopts a common set of weighting rules for common keywords, either constant (i.e., global) throughout the system, or following a specific user group profile, or assigning discrete weights following a coarse step function with lazy update on the keyword counts, upon document cluster change.

To strike a balance between information amount and size of documents so as to maximize document transmission efficiency among a set of documents, the equivalent notion of RIC can be defined. This is known as *INformation Density (IND)*.

**Definition 8.** The *information density*,  $\Gamma_i$ , of information unit  $n_i$  is the ratio of its absolute information content to its length, i.e.,  $\Gamma_i = P_i/l_i$ .

It can be observed that transmitting information units according to information density order can maximize the perceived information content at a client.

### 3.4. Multi-resolution transmission at work

We illustrate as an example the various notions of information content for the information units contained within a sample manuscript. This is illustrated in Table 1. The abstract is considered as Section 0. Paragraphs not belonging to any subsection are grouped under a virtual subsection for simplicity. For instance, all paragraphs belonging to Section 3, but not to Subsection 3.1, are grouped under the virtual Subsection 3.0. This allows us to create a balanced document-tree. The query-based information content is determined with respect to the query  $Q = \{\textit{browsing, mobile, web}\}$ . Equations, figures and tables are removed before generating the SC. A prototype had been implemented based on some of the notions of information content, for the special case of intra-document flat transmission ( $\mu = \lambda$ ).

The transmission order of the document at section LOD (i.e.,  $\mu = \lambda = 2$ ) based on the notion of PIC is ⟨Section 3, Section 1, Section 2, Section 4, Section 0⟩. Notice that Section 0 is found to contain the least amount of information. This is understandable as Section 0 is the abstract of the paper and it is relatively short, hence the low PIC. Based on RIC, the order would be ⟨Section 1, Section 2, Section 4, Section 3, Section 0⟩. Thus, Introduction and Related Work are detected to be more important and transmitted earlier by using RIC. Here, Abstract does not seem to be well-written, or may be reusing keywords that occur many times to bear a smaller weight, or due to the absence of contextual adjustment. Introduction tends to contain more new terms that occur only a few times, thus bringing up the RIC a bit. If subsection LOD is preferred (i.e.,  $\mu = \lambda = 3$ ), some subsections of Section 3 still contain more information than other subsections under some other sections. The transmission is in the order of ⟨Subsection 3.2, Subsection 3.0, Subsection 3.1, Subsection 1.0, Subsection 2.0, Subsection 4.0,

Table 1  
Information content of a sample manuscript

Sect./Subsect./Para.	PIC $p$	RIC $\gamma$	QIC $q^Q$	MQIC $\tilde{q}^Q$	RQIC $\gamma^Q$	RMQIC $\tilde{\gamma}^Q$
0	0.04891	$0.45371 \times 10^{-4}$	0.15119	0.13635	$0.14025 \times 10^{-4}$	$1.26484 \times 10^{-4}$
0.0	0.04891	$0.45371 \times 10^{-4}$	0.15119	0.13635	$0.14025 \times 10^{-4}$	$1.26484 \times 10^{-4}$
0.0.0	0.04891	$0.45371 \times 10^{-4}$	0.15119	0.13635	$0.14025 \times 10^{-4}$	$1.26484 \times 10^{-4}$
1	0.11773	$0.53807 \times 10^{-4}$	0.33231	0.30771	$0.15188 \times 10^{-4}$	$1.40635 \times 10^{-4}$
1.0	0.11773	$0.53807 \times 10^{-4}$	0.33231	0.30771	$0.15188 \times 10^{-4}$	$1.40635 \times 10^{-4}$
1.0.1	0.03924	$0.73209 \times 10^{-4}$	0.18087	0.16784	$0.33744 \times 10^{-4}$	$3.13134 \times 10^{-4}$
1.0.2	0.04899	$0.47749 \times 10^{-4}$	0.12543	0.11461	$0.12225 \times 10^{-4}$	$1.11706 \times 10^{-4}$
1.0.3	0.01327	$0.46076 \times 10^{-4}$	0.00000	0.00158	$0.00000 \times 10^{-4}$	$0.05486 \times 10^{-4}$
1.0.4	0.01623	$0.48018 \times 10^{-4}$	0.02601	0.02369	$0.07695 \times 10^{-4}$	$0.70089 \times 10^{-4}$
2	0.11619	$0.50517 \times 10^{-4}$	0.09738	0.10083	$0.04234 \times 10^{-4}$	$0.43839 \times 10^{-4}$
2.0	0.00619	$0.50517 \times 10^{-4}$	0.09738	0.10083	$0.04234 \times 10^{-4}$	$0.43839 \times 10^{-4}$
2.0.1	0.05327	$0.44878 \times 10^{-4}$	0.03402	0.03897	$0.02866 \times 10^{-4}$	$0.32831 \times 10^{-4}$
2.0.2	0.01404	$0.44151 \times 10^{-4}$	0.13131	0.01256	$0.41292 \times 10^{-4}$	$0.39497 \times 10^{-4}$
2.0.3	0.04888	$0.61484 \times 10^{-4}$	0.05023	0.04932	$0.06358 \times 10^{-4}$	$0.62038 \times 10^{-4}$
3	0.60511	$0.47109 \times 10^{-4}$	0.29368	0.33276	$0.02286 \times 10^{-4}$	$0.25906 \times 10^{-4}$
3.0	0.15982	$0.50704 \times 10^{-4}$	0.11589	0.11690	$0.03677 \times 10^{-4}$	$0.37088 \times 10^{-4}$
3.0.1	0.04710	$0.56953 \times 10^{-4}$	0.07853	0.07087	$0.09496 \times 10^{-4}$	$0.85695 \times 10^{-4}$
3.0.2	0.06370	$0.44483 \times 10^{-4}$	0.02422	0.02933	$0.01691 \times 10^{-4}$	$0.20482 \times 10^{-4}$
3.0.3	0.04902	$0.54894 \times 10^{-4}$	0.01313	0.01670	$0.01470 \times 10^{-4}$	$0.18701 \times 10^{-4}$
3.1	0.12141	$0.54615 \times 10^{-4}$	0.17778	0.17737	$0.07997 \times 10^{-4}$	$0.79789 \times 10^{-4}$
3.1.1	0.04894	$0.61871 \times 10^{-4}$	0.15512	0.14700	$0.19611 \times 10^{-4}$	$1.85841 \times 10^{-4}$
3.1.2	0.03451	$0.43138 \times 10^{-4}$	0.02268	0.02586	$0.02835 \times 10^{-4}$	$0.32325 \times 10^{-4}$
3.1.3	0.03796	$0.60063 \times 10^{-4}$	0.00000	0.00451	$0.00000 \times 10^{-4}$	$0.07136 \times 10^{-4}$
3.2	0.19599	$0.43787 \times 10^{-4}$	0.00000	0.02329	$0.00000 \times 10^{-4}$	$0.05203 \times 10^{-4}$
3.2.1	0.03931	$0.47881 \times 10^{-4}$	0.00000	0.00467	$0.00000 \times 10^{-4}$	$0.05688 \times 10^{-4}$
3.2.2	0.05642	$0.40678 \times 10^{-4}$	0.00000	0.00671	$0.00000 \times 10^{-4}$	$0.04838 \times 10^{-4}$
3.2.3	0.00559	$0.44720 \times 10^{-4}$	0.00000	0.00665	$0.00000 \times 10^{-4}$	$0.05312 \times 10^{-4}$
3.2.4	0.04426	$0.43563 \times 10^{-4}$	0.00000	0.00526	$0.00000 \times 10^{-4}$	$0.05177 \times 10^{-4}$
3.3	0.06293	$0.43490 \times 10^{-4}$	0.00000	0.00748	$0.00000 \times 10^{-4}$	$0.05169 \times 10^{-4}$
3.3.1	0.01868	$0.38755 \times 10^{-4}$	0.00000	0.00222	$0.00000 \times 10^{-4}$	$0.04606 \times 10^{-4}$
3.3.2	0.04425	$0.45855 \times 10^{-4}$	0.00000	0.00526	$0.00000 \times 10^{-4}$	$0.05451 \times 10^{-4}$
3.4	0.06496	$0.41991 \times 10^{-4}$	0.00000	0.00772	$0.00000 \times 10^{-4}$	$0.04990 \times 10^{-4}$
3.4.1	0.03552	$0.41064 \times 10^{-4}$	0.00000	0.00422	$0.00000 \times 10^{-4}$	$0.04879 \times 10^{-4}$
3.4.2	0.02943	$0.43152 \times 10^{-4}$	0.00000	0.00350	$0.00000 \times 10^{-4}$	$0.05132 \times 10^{-4}$
4	0.11196	$0.49256 \times 10^{-4}$	0.12543	0.12209	$0.05518 \times 10^{-4}$	$0.53713 \times 10^{-4}$
4.0	0.11196	$0.49256 \times 10^{-4}$	0.12543	0.12209	$0.05518 \times 10^{-4}$	$0.53713 \times 10^{-4}$
4.0.1	0.02393	$0.51909 \times 10^{-4}$	0.06336	0.05724	$0.13744 \times 10^{-4}$	$1.24165 \times 10^{-4}$
4.0.2	0.01897	$0.43811 \times 10^{-4}$	0.05074	0.04577	$0.11718 \times 10^{-4}$	$1.05704 \times 10^{-4}$
4.0.3	0.04625	$0.51561 \times 10^{-4}$	0.00000	0.00550	$0.00000 \times 10^{-4}$	$0.06132 \times 10^{-4}$
4.0.4	0.02281	$0.47324 \times 10^{-4}$	0.01134	0.01359	$0.02353 \times 10^{-4}$	$0.28195 \times 10^{-4}$

Subsection 3.4, Subsection 3.3, Subsection 0.0), with PIC. With RIC, the order would be (Subsection 3.1, Subsection 1.0, Subsection 3.0, Subsection 2.0, Subsection 4.0, ...)

Using different values of upper and lower LOD limits, there is a more flexible ordering. For example, if we choose  $\mu = 2$  and  $\lambda = 3$ , with PIC, the top sequence still follows the ordering of (Section 3, Section 1, Section 2, Section 4, Section 0); however the subsections are no longer transmitted sequentially, but in the order of PIC: (Subsection 3.2, Subsection 3.0, Subsection 3.1, Subsection 3.4, Subsection 3.3, Subsection 1.0, Subsection 2.0, Subsection 4.0, Subsection 0.0). If we choose  $\mu = 2$  and  $\lambda = 4$ , the top sequence

is unchanged, but there is yet further flexibility. With PIC, we have (Subsubsection 3.2.2, Subsubsection 3.2.3, Subsubsection 3.2.4, Subsubsection 3.2.1, Subsubsection 3.0.2, Subsubsection 3.0.3, Subsubsection 3.0.1, Subsubsection 3.1.1, Subsubsection 3.1.3, Subsubsection 3.1.2, Subsubsection 3.4.1, Subsubsection 3.4.2, Subsubsection 3.3.2, Subsubsection 3.3.1, . . . , Subsubsection 4.0.3, Subsubsection 4.0.1, Subsubsection 4.0.4, Subsubsection 4.0.2, Subsubsection 0.0.0).

We conclude this example with  $\mu = 3$  and  $\lambda = 4$  based on PIC. The transmission order is (Subsubsection 3.2.2, Subsubsection 3.2.3, Subsubsection 3.2.4, Subsubsection 3.2.1, Subsubsection 3.0.2, Subsubsection 3.0.3, Subsubsection 3.0.1, Subsubsection 3.1.1, Subsubsection 3.1.3, Subsubsection 3.1.2, Subsubsection 1.0.2, Subsubsection 1.0.1, Subsubsection 1.0.4, Subsubsection 1.0.3, Subsubsection 2.0.1, Subsubsection 2.0.3, Subsubsection 2.0.2, . . . , Subsubsection 3.4.1, Subsubsection 3.4.2, Subsubsection 3.3.2, Subsubsection 3.3.1, Subsubsection 0.0.0).

#### 4. Conclusion

We have presented an integrated framework for multi-resolution transmission of documents and collection of documents to mobile clients. Intra-document multi-resolution transmission allows mobile clients to visualize documents at any selected level of detail. Based on the different notions of information content, clients are presented with the main document content before supplementary information in a progressive manner. Furthermore, transmission granularity can be varied through proper definition of the upper LOD limit and lower LOD limit. Information content and its various transformations are integrated and explained, addressing different transmission constraints. This is important in a mobile environment where bandwidth is scarce. On the other hand, inter-document multi-resolution transmission allows clients to perceive a high-level view of not only a document, but a collection of related documents, in form of a document cluster graph. The graph can be visualized and the important high-level information units for each constituent document can be prefetched to the clients, following an order determined by the information content of the documents and their representative units.

While it has been shown that the use of relative information content in ordering information units for transmission can yield optimal performance in terms of user perceived information over time, it does not take into account the granularity and more importantly, the integral nature of the units. A better cost function and hence notion of information content and its application is necessary in order to produce a best performance. Furthermore, the best information unit for prefetching at document level is yet to be explored. The algorithms adopted in this paper only represent the initial efforts. Finally, the tradeoff between prefetching of high-level information units against transmission of information units within the currently interested document needs to be studied, when there is only one single channel. Alternatively, this can be modeled by the dynamic bandwidth allocation problem between the two logical channels for intra-document transmission and inter-document transmission.

#### References

- [1] M. Ackerman, D. Billsus, S. Gaffney, S. Hettich, G. Khoo, D.J. Kim, R. Klefstad, C. Lowe, A. Ludeman, J. Muramatsu, K. Omori, M. Pazzani, D. Semler, B. Starr and P. Yap, Learning Probabilistic User Profiles, *AI Magazine* **18**(2) (1997), 47–56,
- [2] R. Alonso and H. Korth, *Database System Issues in Nomadic Computing*, In Proceedings of the ACM SIGMOD International Conference on Management of Data, May, 1993, pp. 388–392.
- [3] R. Armstrong, D. Freitag, T. Joachims and T. Mitchell, *WebWatcher: A Learning Apprentice for the World Wide Web*, 1995.

- [4] N. Belkin and W.B. Croft, Information Filtering and Information Retrieval: Two Sides of the Same Coin? *Communications of the ACM* **35**(12) (December, 1992), 29–38.
- [5] B.Y.L. Chan, A. Si and H.V. Leong, *Cache Management for Mobile Databases: Design and Evaluation*, In Proceedings of the 14th International Conference on Data Engineering, February, 1998, pp. 54–63.
- [6] D.L. Chan, R.W.P. Luk, W.K. Mak, H.V. Leong, E.K.S. Ho and Q. Lu, *Multiple Related Document Summary and Navigation using Concept Hierarchies for Mobile Clients*, In Proceedings of the 17th ACM Symposium on Applied Computing, March, 2002, pp. 627–632.
- [7] J.M. Follin, A. Bouju, F. Bertrand and P. Boursier, *Management of Multi-Resolution Data in a Mobile Spatial Information Visualization System*, In Proceedings of the Web Information Systems Engineering Workshops, 2003, pp. 92–99.
- [8] S. Geffner, D. Agrawal, A. El Abbadi, T. Smith and M. Larsgaard, *Smart Indexes for Efficient Browsing of Library Collections*, In Proceedings of the IEEE Advances in Digital Libraries Conference, April, 1998, pp. 107–116.
- [9] K.K. Hoi and D.L. Lee, *Document Visualization on Small Displays*, In Proceedings of the International Conference on Mobile Data Management, 2003, pp. 262–278.
- [10] T. Imielinski and B. Badrinath, Mobile Wireless Computing: Challenges in Data Management, *Communications of the ACM* **37**(10) (October, 1994), 18–28.
- [11] H.V. Leong, *Browsing Document Clusters on Mobile Web*, In Proceedings of 1999 ACM Digital Library Workshop on Organizing Web Space, August, 1999, pp. 76–90.
- [12] H.V. Leong, D. McLeod, A. Si and S.M.T. Yau, *On Supporting Weakly-Connected Browsing in a Mobile Web Environment*, In the 20th International Conference on Distributed Computing Systems, 2000, pp. 538–546.
- [13] W.S. Li, K.S. Candan, Q. Vu and D. Agrawal, *Retrieving and Organizing Web Pages by Information Unit*, In Proceedings of Tenth International World Wide Web Conference, May 2001, pp. 230–244.
- [14] J.Y.S. Mak, H.V. Leong and A.T.S. Chan, *Dynamic Structuring of Web Information for Access Visualization*, In Proceedings of the 17th ACM Symposium on Applied Computing, March, 2002, pp. 778–784.
- [15] I. Mani and E. Bloedorn, *Maching Learning of Generic and User-Focused Summarization*, In Proceedings of AAI, July, 1998, pp. 821–826.
- [16] M.L. Mauldin and J.R. Leavitt, *Web-Agent Related Research at the CMT*, In Proceedings of ACM International Conference on Networked Information Discovery and Retrieval, 1994.
- [17] T. Parker, Mobile Wireless Internet Technology Faces Hurdles, *IEEE Computer* (March, 1998), 12–14.
- [18] B. Pinkerton, *Finding What People Want: Experiences with the WebCrawler* In Proceedings of Second International WWW Conference: Mosaic and the Web, 1994.
- [19] M. Weiser, Some Computer Science Issues in Ubiquitous Computing, *Communications of the ACM* **36**(7) (July, 1993), 74–84.
- [20] E.Y.C. Wong, A.T.S. Chan and H.V. Leong, *Efficient Management of XML Contents Over Wireless Environment by Xstream*, In Proceedings of the 19th ACM Symposium on Applied Computing, March, 2004, pp. 1122–1127.
- [21] C.C. Yang and F.L. Wang, *Fractal Summarization for Mobile Devices to Access Large Documents on the Web*, In Proceedings of the International World Wide Web Conference, 2003, pp. 215–224.
- [22] S.M.T. Yau, H.V. Leong, D. McLeod and A. Si, On Multi-Resolution Document Transmission in Mobile Web, *SIGMOD Record* **28**(3) (1999), 37–42.

---

**Hong Va Leong** received his PhD from the University of California at Santa Barbara, and is currently an associate professor at the Hong Kong Polytechnic University. He has served as the co-chairs and members in the program committees and organization committees for many international conferences. He is also a reviewer for a number of international journals, including ACM Transactions, IEEE Transactions and Information Systems. His research interests are in mobile computing, internet computing, distributed systems, distributed databases, and digital libraries. He is a member of the ACM and IEEE Computer Society.

**Antonio Si** received his PhD from the University of Southern California, and is currently with Oracle Corporation. Before joining Oracle, he has worked as an assistant professor at the Hong Kong Polytechnic University and as a software engineer at Sun Microsystems Inc. He has served on the program committees for several international conferences and as external reviewers for a number of international conferences and journals. His research interests are in mobile computing, internet computing, and digital libraries. He is a member of the ACM and IEEE Computer Society.





**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

