



Information Overload

Inaccessible data and a
knowledge management
solution

Inaccessible project documents can seriously hamper the capabilities of development teams. This paper investigates the breadth and depth of this problem, describes its consequences, and makes recommendations to resolve document accessibility issues.

August 2002

Table of Contents

Introduction	2
Background	2
Problem Definition	2
Finding the right document	2
Finding the right information in a document	2
Investigative Approaches	3
Evolutionary approach	3
Semi-structured interview approach	3
Intel product lifecycle	3
Problem Identification	3
Causal factors	3
Impact	4
Findings	5
User segmentation: number one priority	5
User segmentation: initial results	5
Technical expertise segment	5
Support and environment segment	5
User segmentation: proposal for discovery	6
Recommendations	6
Understanding the users	6
Finding the right document	6
Cultural aspect: discipline	6
Initial metadata candidate set	6
Finding the right information in a document	6
Conclusions	7
Appendix A: Bibliography	7
Appendix B: Interview Questions	7

Introduction

Information overload can severely hinder the effectiveness of development teams. Important data found in project documents may be inaccessible because the documents are stored in numerous places under various standards. The documents themselves may impose additional barriers if they are large and without navigational aids and executive summaries. In today's fast-paced, information-saturated environment, few people have time for extensive searching to locate the data they need. Instead, they tend to establish local databases or find other ways of getting information. This lack of a cohesive system to make project data accessible results in wastage of resources such as paper, disk space, and, most importantly, time.

This paper describes how Intel IT determined where an information overload problem existed, identified whom the problem impacted, and defined how it impacted them. We also recommend short- and long-term solutions to alleviate this information overload problem.

Background

For our investigation, we interviewed document users at different organizational levels and in different job functions. We used a semi-structured interview format that allowed maximum freedom of expression for the interviewees, while ensuring we received the information we needed within our very restrictive timelines.

We discovered that the problem was more severe than initially expected. Difficulties extended beyond the time spent searching for the right version of the right document, which could take several hours for a single document. Time was also consumed in rework brought about by ambiguous or changing requirements. Additionally, resources such as disk space were misspent through the duplication of documents. This duplication created confusion as well; it was not clear which version was the most recent.

Problem Definition

Finding the right document

The interviewees experienced problems in gaining access to project documents. The process of finding a document often involved locating the document's author and then requesting the document directly from that person. Existing documents were stored in different locations and their version status was often unclear.

Intel IT set out to understand the problem facing those users and to propose an initial solution.

Finding the right information in a document

One of the challenges the interviewees spoke of was finding a specific piece of information within a document. This task consumed a significant amount of time, especially for the longer documents. Documents did not always contain summaries or indexing methods to guide users to the desired information. The sheer size of some documents, at times over 100 pages, became prohibitive. This problem became particularly frustrating when the desired information was a summary that the document did not contain or provide. One interviewee

reported that he often looked for an executive summary, could not find it, and did not have the hours necessary to read the entire document. Instead, he resorted to getting this kind of information from meetings.

Investigative Approaches

Evolutionary approach

Our general approach to the problem is evolutionary and ongoing: We plan to develop the solution iteratively, with the research described here as step one of the solution. We also limited the scope of our initial exploration to include only the project lifecycle documents.

Intel product lifecycle

The Intel Product Lifecycle is a high-level framework process that is used to take a product from market research through production and eventually to product discontinuance.

The use of a product lifecycle is a concept proven by premier companies to:

- Increase development efficiency
- Accelerate time to money (TTM)
- Provide synchronized and stable roadmaps

Semi-structured interview approach

The goal of the interviews was to gain a better understanding of the problems facing real users of the project lifecycle documents, as well as to define a suitable metadata set. A semi-structured interview format proved the best approach for several reasons. For the initial fact-finding exploration, we wanted users to lead the discussion in the direction they felt was most appropriate, towards the area where they experienced the most pain. We therefore chose to include the Critical Incident Technique in the interviews, which ruled out the structured interview format. However, the tight deadline for our investigation demanded complete responses to our questions, which dictated the use of some structured questions. The semi-structured format provided benefits from both approaches. It was open enough to allow free format while simultaneously ensuring that we received the responses necessary to meeting our goals.

Refer to Appendix B for the interview questions. The interviewer used the bolded questions to lead the interview, and the plain text and bulleted questions to clarify answers and draw out the interviewees where necessary.

The findings of this paper are based on interviews of six Intel employees. Though tight deadlines restricted the number of interviews, six proved sufficient for understanding the problem and making recommendations for both the next steps and the metadata candidate set. The most significant drawback was that the six people interviewed did not completely represent all of the different user groups. However, this initial investigation allowed us to conclude that user segmentation must be included in the next research phase to address the very divergent needs of Intel employees.

Problem Identification

Causal factors

The number one inhibitor to finding documents, and the most common factor reported among the interviewees, was that no consistent process existed for storing and managing the documents. Documents were stored in a number of different repositories, each using different storage mechanisms. Those repositories included:

- Group calendaring systems
- Local Web sites
- Shared drives
- Local hard disks
- Document management systems

Different groups used these repositories in different fashions and without a consistent process for depositing documents in any of the repositories. This problem was particularly severe for globally dispersed teams (GDTs). Teams at different sites often used different repositories; they had varying degrees of established processes for depositing documents, and they followed these processes to varying degrees. For example, one team had sub-teams on both the east and west coasts of the United States. The west coast sub-team's preferred method of posting documents was on its local Web site, whereas the east coast sub-team used the group calendaring system.

The users confronted diverse problems. A significant problem was the lack of communication regarding the repositories: where the documents were stored and

what other document repositories existed. Even when the users knew which document repositories existed, access to them was often problematic, since their current locations were not known. This situation manifested itself frequently and under different circumstances. For example, one user knew a certain project had a Web site but did not know how to gain access to it. Another common problem was with the location of shared drives and how to gain access to them. Different divisions had their own Web sites and their own processes of updating those Web sites, with no standard across the different organizations. The problem was compounded since almost every project team handled documents in a unique manner and with no consistent standards across projects, products, or organizations. Reorganizations brought more disarray to how documents were stored within groups and projects.

Another aspect of the document storage problem was that the information was not well archived with proper revision controls, resulting in the original version of a document often being inaccessible and sometimes nonexistent. Users rarely knew the location of the latest revision and were often unclear which document was the latest version. Because no standard methodology existed for document posting, some documents were not posted at all. For example, in a certain project a document was reviewed, approved, and emailed to over 50 people, but it was never posted to any public repository.

Users had to find a contact involved in the project if they could not find the needed information (which could be the correct version of a document or a specific piece of data from a document) in any of the known document repositories. The best such contact was the project manager, or the engineer responsible for assisting customers with designs or marketing. Finding these contacts was not always an easy task. They tended to be busy and often had full schedules. Moreover, users often needed to contact several people to find the information they were looking for. One interviewee mentioned that he frequently contacted three to five people before finding the information he sought.

Certain specific groups or projects had reached an agreement within their respective teams on an established process and methodology for posting documents. The people directly involved in the project were kept aware of changes thanks to their ongoing involvement in the project. People outside the group, who may have

had a legitimate interest and reason to find information related to the project (such as high-level managers, marketing people, or new employees) but were not heavily involved with it, encountered the same problems found in groups that had not agreed on established processes.

Another problem was that, once a user found the right document and gained access to it, the document could be too lengthy to be useful. One interviewee mentioned that he often came across documents exceeding 100 pages and did not have time to thoroughly read them. This situation created an urgent need for access to summaries and high-level overviews, which might have been all that the user required. Such summaries, however, did not exist for most documents.

Impact

The inability to gain access to the right information often had diverse negative effects. The most obvious negative effect was resource wastage for paper, disk space, and, most notably, time. One interviewee estimated 15% to 20% rework due to lack of clear and current information. Rework stemmed from one of two causes: (1) developing to requirements that had changed but were not promulgated to everyone involved, or (2) developing to ambiguous requirements that were interpreted inconsistently. In general, people tended to share information only at its end state, when it was ready for consumption, and not during discovery. This situation also created duplication of efforts: Several groups re-discovered known problems that had not been communicated to others. As a result, they implemented localized solutions, such as a database for known software defects that contained information regarding how to reproduce the defect, who owned the defect, what its priority was, and so forth.

Searching for the right information also resulted in a significant loss of time. Some users estimated they spent several hours, spread over several days, locating a single document. Other estimates were 10 to 15 minutes on average spent looking for a single document.

The difficulty in accessing the right information created a new behavior trend for some users: They sought out information in meetings. One interviewee reported that he found information exclusively through meetings and personal interactions with others, completely circumventing the need for accessing the project documents

himself. More than one interviewee listed meetings as an important source for information. Key programs had meetings frequently enough to satisfy the information needs of another interviewee. The meetings also provided a forum for online questions and answers, an interaction that was missed with documents alone. The documents from these meetings were used as reference material and were often stored on personal hard drives (another waste of disk space).

When documents were not easily accessible, users could get only a snapshot of the environment *unless* they knew whom to ask. To resolve this problem, specific groups or projects established unique processes to address this problem, as in the examples above.

Findings

User segmentation: number one priority

The most striking finding was that the interviewees had widely divergent views on how project documents are accessed and used by different users. The needs of users from different job functions varied greatly in terms of (1) role or job function, (2) experience, and (3) geographical location. At one extreme, users needed only high-level (highlight) information: an executive summary of the project's goals, its timelines, or perhaps even an easy access to feature sets. Users often acquired this kind of information from meetings. At the other extreme, such as when a current project relied heavily on a previous revision, users needed very detailed information. In this last instance, users needed historical data to verify what the original decisions were and to understand how they were made. Such information was also very useful in the development of best known methods (BKMs) and baseline assumptions.

Interviewees rarely had an inclusive picture of the different ways the project documents were used. One interviewee from marketing did not use the documents and exclusively sought the information he required through interactions with people. Another interviewee did not see the need for a solution beyond the location of all the documents in a single repository with a predefined navigational structure. This interviewee, from the technical domain, stated that a single repository with an easy navigational structure should be sufficient for everyone, with no need for search capabilities. Indeed, this solution proved sufficient for the dozens of people directly involved in that specific project. Previously,

users had experienced great confusion as to where the documents were stored. After this group began using a single process for depositing its documents, the accessibility problem decreased significantly, though it did not disappear completely. Documents were still not always posted in the prescribed places. Some were posted instead to Web sites or shared drives (with the added complications of access rights).

The developers directly involved in the project could sort through this information maze, but users who were not fully involved in the project and new members coming on board had difficulty. Users not directly involved in the project were often unfamiliar with the group's established process.

This problem became even more challenging after reorganizations occurred, or projects and/or products underwent name changes. It grew more severe as the number of products and projects increased.

User segmentation: initial results

The results from this initial round of interviews identified two main user segments. Since the two segments have different needs, the members of these different segments look for information in different ways and may have to be addressed separately.

Technical expertise segment

The technical expertise segment is comprised of engineers, developers, sustainers, and liaisons to the systems that were developed. This group normally works very closely with a single project at a time and has a very thorough understanding of the project's details and processes. Users in this group might need extremely detailed information (for example, when working on a new version of an existing product).

Support and environment segment

The support and environment segment encompasses several distinct job functions:

- Marketing segment
- High-level management
- New (to Intel or to the project) technical employees

This segment often requires less detailed information about the project, for example, summaries of the project objectives, status, timelines, product overview and major features, and so forth.

User segmentation: proposal for discovery

We needed to explore user segments along three known dimensions:

- Role or job function
- Prior experience
- Geographical location

Users identified all three dimensions as very important. This paper briefly discussed most of the aspects of these dimensions, but one aspect of geographical location has not yet been mentioned: Different cultures often have different approaches and processes for tackling problems. These differences were clearly evident in globally dispersed cultures, and examples abound.

Recommendations

Based on our study, we can make various recommendations to increase team confidence that the product delivered is the product agreed upon (through consistency and accessibility of documentation) and to eliminate ambiguity and resource wastage as the group implements the recommendations.

Understanding the users

Every interviewee in this study was located in the United States, whereas proper user segmentation must include users from different geographies. Additional interviewees must be selected according to their experience, particularly with respect to their tenure at Intel, to their job functions, and to their scope of managerial responsibilities.

To provide a solution that will benefit all users, it is important to understand the needs of all the different population segments, hence the need for user segmentation.

Finding the right document

Finding the right document in the current environment was the biggest challenge identified during the interviews. The first short-term recommendation is to implement a single document repository that includes revision control (a critical ingredient for some interviewees).

Cultural aspect: discipline

Once a group makes a single repository available with clear requirements, expectations, and processes in place for the posting of documents, the success of this

solution depends on discipline. Discipline is required in both the adoption of and the adherence to the new processes. Without such discipline, this solution will not be successful.

Initial metadata candidate set

The second short-term recommendation in this area is the initial metadata candidate set, by which users can easily gain access to the correct document.

The initial metadata candidate is based on the interviews. The interviewees had many diverse needs, some of which were almost directly opposite to others. For example, one interviewee found it critical to have only the latest version of any document, while another found it extremely important to access the first version of the document. Still another did not find either of these features important. This situation led us to conclude that, to identify different needs, we must conduct user segmentation. Therefore, the initial metadata candidate set does not address all these needs. Instead, we include only the few metadata tags that will provide obvious value to most of our interviewees.

Whenever possible, the metadata derives its values from controlled vocabularies. This method helps avoid the possibility of typos and other mistakes. Where a controlled vocabulary is not possible (for example, on a date value), a controlled format should be enforced (for example, Jun 21, 2001: the month's three-letter abbreviation, followed by a two digit day-in-month, followed by a comma, followed by a four digit year).

Finding the right information in a document

Many possible solutions exist for the problem of finding the right information in a document. A solution that would satisfy the interviewees, and was a critical aspect for some of them, is the inclusion of an executive summary in all project documents. Currently, according to the interviewees, most of the documents do not contain executive summaries.

We examined two other proposals as well. The first stems from the emerging field of research called knowledge discovery in databases (KDD), also commonly referred to as data mining. This field tackles the problem of "identifying valid, novel, potentially useful, and ultimately understandable patterns in data" (refer to Appendix A). Some branches of this field examine textual data and could therefore be well suited to extracting summaries from project lifecycle documents.

Our alternate proposal involves requirement tagging. Several interviewees wanted the ability to follow specific requirements throughout the project documents. We propose a metadata tag that contains a unique identifier for each requirement. When a user requests a specific, unique requirement identifier tag for a specific project, the metadata tag would present all the references, and the context of the associated requirements from the relevant project documents, to the user.

Conclusions

The information overload problem inhibits users' abilities to easily find and use information contained in the project lifecycle documents for two main reasons:

- Users do not know where to find the appropriate revision of the required document and how to gain access to it.
- Users often find it difficult to extract the necessary information from a document after it has been located.

An initial, short-term solution to these challenges is a single, centralized repository with revision control. The repository hosts all project documents and makes them accessible via the recommended candidate metadata set. This solution makes documents easily accessible and assures users that the documents are accurate and current. The repository should be implemented with the recommended candidate metadata to enable fast access through commonly requested document attributes.

This solution is a first step and must be refined.

For any long-term solution, the number one priority is identifying user segmentation. The user population is widely divergent, with different population segments having vastly different needs and requirements. Project members and project "outsiders" (for example, technical engineers versus marketing personnel and high-level managers) need to gain access to and acquire project information in significantly different ways. User segmentation enables us to identify and respond to these various, specific needs and is a critical long-term solution component.

Appendix A: Bibliography

[FPSS96] U.M. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. *Advances in Knowledge Discovery and Data Mining, Chapter 1: From Data Mining to Knowledge Discovery: An Overview*, pages 1—34, AAAI Press, 1996.

Appendix B: Interview Questions

(The bolded questions were used to lead the interview. The plain text and bulleted questions were used to clarify answers and draw out users where necessary.)

Tell me about yourself, what you do [Who is the user?]

Job function of the information-seeker?

What is the nature of the work done by your team?

Recall/explain a recent situation where you tried to find a project document and could not find it? [Critical incident technique]

Why are you looking for the project documents?

What kind of task prompts the project document search?

What is the goal of the project document information search?

Are the documents you are looking for written for some other purpose than the one you are searching them for?

- What was the original purpose of the documents?

What kind of information are you looking for?

Which ones of the project documents do you need to access most?

What specific type of information in the project document are you looking for?

- Updates on product/project?
- Information on features? Functions? Qualities?

What type of project documents do you search for most frequently?

- Prioritize/rate

What type of (prioritize/rate) information in the project documents do you search for most frequently?

How important is it that the project document you retrieve is the most recent one?

Is it important to have historic data (product in pre-production as opposed to production may change names)?

Do you have a need to locate archived, historic or superseded versions of the documents?

How frequently does the information in the project documents change? Does it matter to the results?

How are you looking for the information now?

How would you search for the documents now?

What types of terms do you use?

- What families of terms?

What structures/hierarchies exist now?

Is the information in the project document you get now "the right" information (what you were looking for)?

Is the speed in which you get the project document you're looking for satisfactory?

How long does it take you to find the right document currently? (For success measures)

What would make the experience better?

Why do you think you're having problems finding the information?

Why do you need a better/different way of finding the information?

What would you like to get that you're not getting now?

What do you think can be improved to make the information-accessing better?

Describe some of the major challenges.

How would rate the importance or impact of the things that did not work?

What do you think were the roadblocks?

- Did not have right permissions?
- Information on someone else's hard drive?
- Titles not reflecting information?
- Information not up to date?

Do you find information that's not relevant?

- Why is the information not relevant?
- Updates not available?



For more information

To learn more about IT@Intel, visit our site on the World Wide Web at www.intel.com/IT.

This document and related materials and information are provided "as is" with no warranties, express or implied, including but not limited to any implied warranty of merchantability, fitness for a particular purpose, non-infringement of intellectual property rights, or any warranty otherwise arising out of any proposal, specification, or sample. Intel assumes no responsibility for any errors contained in this document and has no liabilities or obligations for any damages arising from or in connection with the use of this document.

Intel and the Intel Logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.