REFERENCES

A


B


References


C


References


D


E


References


F


References


References

**H**


**I**


**J**


**K**

Kang, B. Word document and powerpoint slides at:


L


M


Context Proceedings of the 6th Int. Conf. on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems Edinburgh.

N


O


P


References


References


T


APPENDIX A – TROUBLE-SHOOTING SURVEY

This Appendix presents the results of the FastFIX survey at HTG’s Sydney-based support group, undertaken from March through to May 2005.

The survey results include data from:

- 18 respondents in Survey Part A including 9 Technical Support Level 1 (TS1) and 9 Technical Support Level 2 (TS2) respondents, covering both Product Group 1 (PG1) and Product Group 2 (PG2), and
- 20 respondents in Survey Part B\(^{195}\).

The researcher’s interpretations, thoughts, and personal opinions are recorded in this report as Researcher’s Notes.

\(^{195}\) The names of the product groups have been aliased to protect the confidentiality of HTG.
Survey Part A – The Trouble-shooting Environment

Q. Who is solving problems, and how are they doing it?

Question 1: Survey Groups

There were 9 TS1 respondents and 9 TS2 respondents for Survey Part A. Of the 9 TS1 respondents, there were 4 from PG1, and 5 from PG2. Of the 9 TS2 respondents, there were 7 from PG1, and 2 from PG2.

Questions 2 to 7: Experience and Expertise

TS1 respondents had on average 7 years experience with 4 years of experience relevant to their role at HTG. Those working in TS1 PG1 had been with HTG on average for 10 months while those working in TS1 PG2 had been with HTG on average for 2.5 years. As shown in Figure 56 (RHS), the 9 TS1 respondents rated their level of expertise as senior (2), middling (1), junior (5), and novice (1).

TS2 respondents had on average 12 years ICT industry experience, 11 years of experience relevant to their role at HTG, and they had been working at HTG for an average of 4 years. As shown in Figure 56 (LHS), the 9 TS2 respondents rated their level of expertise as expert (4 respondents), senior (2 respondents) and middling (3 respondents).

Figure 56: Self-assessed expertise of TS2 respondents compared to the TS1 respondents.
Question 8: Please list the top 3 activities that occupy you at work from week to week. Please include an estimate of the average weekly time that you spend performing each of these activities.

The activities that respondents indicated as occupying them most from week to week included:

- **trouble-shooting and solving customer problems (80%)**, including:
  - accepting the CaseDB case, reviewing the case notes, working the case
  - speaking with and corresponding with customers / problem definition
  - diagnosis including gathering, reviewing and analysing log files and trace files.
  - fault finding / problem isolation
  - problem recreation / simulation / testing in the lab
  - trying different things to fix the problem
  - finding out who to direct questions to
  - root cause analysis (RCA)
  - resolving the problem, including reading product manuals to the customer!
  - closure
  - **research (23%)**, including:
    - in depth trouble-shooting analysis
    - reading emails, case histories, support manuals and trouble-shooting guide updates
    - looking for prior cases with the same issues / searching SolutionDB to find out the solution
    - using Google and Microsoft resources to investigate / research

- **administrative tasks (13%)**, including:
  - documentating flaws
  - emails
Appendix A – Trouble-shooting Survey

- answering the phone
- updating CaseDB cases with customer or engineering feedback / logging notes
- doing timesheets

- **mentoring (6%)** including:
  - consulting to and assisting level 1 support in the quick resolution of cases.

- **self-development (1%)** including:
  - reading and research.

Researchers Note 1:

Two respondents answered that they spent 15-25% of their time with the customer in problem definition. Knowing what questions to ask appears to be the essence of trouble-shooting. In many cases, when the question is ready - the answer will appear. Questions help both inexperienced and experienced trouble-shooters to determine the class of problem on hand. If some of these questions were provided to the customer up front, for example on the web, it would speed up the problem determination process. Three of the respondents to this question answered that they spent 50% of their time looking for prior cases with the same issues. That's a lot of time searching for the answers to known questions.

Given the turnover (only 2.5 years average longevity in PG2 TS1), having this “problem determination” and "where to search, and what to search for" knowledge stored only in people's heads represents an asset that is lost every time sometime walks out the door. Further, one respondent answered that they spent 15% of their time “reading product manuals to the customer”!

The goal of a continuously learning guided trouble-shooting tool would be to reduce both the time taken to define the problem, and the time taken to find the matching solution. The granularity of linked information would match the depth of knowledge about the problem on hand. For instance, at the highest level, the most general level of information should be provided, and at the lowest level, specific solutions should be provided. This type of solution would be well suited to pushing a structured Q-A trouble-shooting protocol out to the customer, with much-needed and timely links to relevant solutions and product information.
Question 9: Please describe the training that you have had which has best contributed to your expertise in your current job role

Respondents answered:

- Various PG2 training including overview and architecture, installation and trouble-shooting, also self training in the lab by testing different software functionalities.
- Very little - product install, the rest has been on the job.
- Hands on training.
- Basic ProductA training, the rest has all been learnt on the job.
- Technical training: PG2 installation trouble-shooting; basic open fibre; basic open system host training; basic ProductA training; customer handling training: customer service / trouble-shooting techniques.
- PG1 management and operations course
- PG1 administrator training
- HTG training, PG1 operations.
- PG1 operations and management; PG1 (version XX) training; ProductB install and configuration; ProductA training; BOS 4, BOF etc.
- Bootcamp, introduction to related background and industry knowledge.
- ProductB manager course; PG1 (version YY) update training; on the job training. On the job training has been the most instructive. The first course was excellent as an introduction to how to use the equipment but not to fault finding.
- External courses in windows / linux / networking, also some internal training.
- PQR training
- Formal training - basic; practical field - different types of situations and usages; practical engineering - having time with specific tasks to do hands on and understanding.
- New product training and engineering training.
Network trouble-shooting given out by a PG1 colleague. The training provided me with the skills to troubleshoot different type of network problems.

ProductA training; ProductB administration training.

At uni - having to use your initiative to find information and learn rather than wait for it to be spoon-fed.

**Question 10: What attributes would you say makes a person good at trouble-shooting and solving customer problems at HTG?**

Respondents answered as follows. For easier interpretation, these answers have been summarised by the author on the following page.

- being calm and composed, logical, open minded, and a good listener
- able to identify and fix customers problems in a timely manner
- understanding customer symptoms / issues
- a thorough understanding of the key features in a product; knowing everything about something and knowing something about everything
- knowledge of the product; confidence; *know the questions to ask*
- ability to think laterally, large skill base and ability to learn
- a combination or logical and creative thinking, analytical thinking, and empathy - to really help the customer.
- logical, congenial
- perserverance, lateral thinking, attention to detail
- logical thinking; a sense of humour - personable; knowledge of the product
- good listener, technical knowledge, ability to use tools effectively, ability to think outside the box, ability to remain calm when user serious pressure, ability to stand your ground.
- abstract thinking; ability to follow a logical argument; understanding that there is a relation between cause and effect
Appendix A – Trouble-shooting Survey

- understanding issues over the phone; ability to work alone; a solid understanding of the way a product should work
- good understanding of overall computer systems; programming or scripting ability; personality - trying to dig into system internals
- willingness to help, knowledge sharing with the others, and document the technical detail for future reference
- communication, listening, technical expertise; intelligence
- doesn't quit, not easily bored, likes a sense of achievement at solving problems

The following additional comment was provided:

- using a database such as SolutionDB helps to share the information if the same kind of issue happened previously and lets everyone know about the solution

Researchers Note 2: The FastFIX architecture is designed specifically to address one of the fundamental secrets of trouble-shooting identified here: “know the questions to ask”.

These responses have been summarised as follows:

Attributes Required of Trouble-shooters

- **problem solving skills (15 responses):**
  - logical / analytical / cause and effect thinking (8 responses)
  - creative / lateral / abstract / outside the box thinking (5 responses)
  - able to identify and fix customers problems in a timely manner
  - know the questions to ask

- **personal attributes (13 responses):**
  - being calm and composed, even when under serious pressure (2 responses)
  - perserverance / doesn't quit / will dig into the system internals (3 responses)
  - large skill base
  - open minded
Appendix A – Trouble-shooting Survey

- ability to learn
- attention to detail
- ability to work alone
- intelligence
- not easily bored
- likes a sense of achievement at solving problems

- **technical knowledge and expertise (10 responses):**
  - knowledge of the product, including a solid understanding of the way a product should work, and a thorough understanding of the key features (4 responses)
  - technical knowledge and expertise (2 responses)
  - knowing everything about something and knowing something about everything
  - ability to use tools effectively
  - good understanding of overall computer systems
  - programming or scripting ability

- **communication skills (9 responses):**
  - good listener / empathy (3 responses)
  - understanding customer symptoms / issues including understanding them over the phone (2 responses)
  - confidence / ability to stand your ground (2 responses)
  - a sense of humour – personable
  - congenial

- **collaboration skills (1 response):**
  - willingness to share knowledge with others, for example by documenting the technical detail for future reference, and to help others
Researchers Note 3: Given the very strong dependency on team-mates when solving customer problems as shown in Q12 there appears to be misalignment between the need and the importance that the team places on collaboration. The need for collaborative skill is a lot higher than respondents have identified. Perhaps the current environment, for example the time pressures, metrics, and incentive structures discourage the team from collaborating? Note: collaboration comprises both the act of mentoring, and the act of asking for help as required.

**Question 11: What is your personal strategy for trouble-shooting technical issues and / or solving customer problems at HTG?**

Respondents described the following processes. For easier interpretation, these answers have been categorised and organised by the author on the following page into a guided trouble-shooting sequence.

- Find out what the problem is by talking to the customer. Check what processes are involved and their current status. Correct the status if incorrect and re-run the customer process to check if it now works. This may include reviewing logs and discussions with the level 2 team as to how to correct issues; also checking the errors in SolutionDB and the Product Support Matrix and related documents.

- 1. understanding what the problem is; 2. try one or two things to fix the problem while on the phone to the customer; 3. gather log files if the problem is not resolved after the first contact; 4. research log files for possible error; 5. research / problem recreation in the lab; 6. if the lab result is showing the same thing then most likely the customer’s problem is a bug; 7. escalate to TS2 for opening a ticket with engineering.

- List and understand the customer’s problem; collect logs if necessary; check HTG documents if necessary (Docco); check SolutionDB knowledge if necessary; check PG2 service alerts if necessary.

- 1. experience (past cases); 2. scan logs; 3. restart / initialise parameters; 4. knowledge database (SolutionDB); 5. colleagues; 6. TS2.

- Narrow the problem; get relevant information; ask the right questions.
• Search SolutionDB, remedy, Docc. *Ask others in the team.* Consult with TS2.

• 1. understand the problem and its impact, 2. find a solution, 3. replicate the problem, 4. test the solution, 5. apply the solution, 6. get feedback about the success of the solution, 7. repeat from step 1 if needed.

• keep the customer informed; troubleshoot thoroughly; investigate and research where possible

• *identify the problem area:* find the error messages; analyse the error messages; identify the piece of code causing the errors; test and verify your findings in the lab; *search the trouble-shooting database for similar symptoms*

• gather as much information as possible from the customer; research known issues that may be related; never take what the customer says as gospel; try to replicate the problem

• rely on knowledge and experience, use all available tools as best I can, attack problems on multiple fronts if required, use desktop Google to search through all previous correspondence as I have everything on my computer, engage the HSL / gns / open sys resources

• I look at the observations collected, I theorize about what could cause the observed behaviour, I test each explanation and use a *process of elimination by finding counter examples* until only one possible explanation remains.

• 1. listen to the customer problem definition; 2. *ask specific probing questions* from above; 3. write own problem definition; 4. view logs - look for error messages; 5. find error message definition; 6. track how message was created

• understand the external symptoms clearly; think about any possible internal causes; look into the logs and messages; analyse the trace if necessary; recreate the problem

• understand the customer concerns and impact; collect and evaluate the information provided; get in touch with other groups for assistance as needed; capture the network trace; notify manager as needed

• listen to the customer problems and understand them; talk to people, look in SolutionDB, read manuals, search the net etc; determine if it is a product or procedure
problem; if it’s a product problem then implement a work around and involve engineering; if it’s a procedural problem involve the field and educate the customer

- 1. review the case including the past efforts of the customer and HTG, 2. review the evidence (logs from the customer), 3. speak with the customer to confirm understanding, 4. investigate using knowledge bases and documentation, 5. suggest a fix, 6. check the outcome with the customer, 7. request new logs from the customer and repeat as necessary from 2.

Researchers Note 4: Many steps in this process involve the collection and interpretation of information. FastFIX aims to map out the path to information in a manner that will make relevant information accessible in the context of the problem on hand. It could also be used to assist with the auto-interpretation of log files.

**Q. Is trouble-shooting a team effort, or an individual effort?**

*Question 12: What percentage of customer problems assigned to you would you solve yourself, without involving other people?*

Responses to this question are shown in Figure 57. On average, respondents said that 57% of problems assigned to them were solved without involving other people. A corollary is that they would involve others in 43% of the problem cases assigned to them.
Responses from TS2 respondents to this question are shown in Figure 58. On average, TS2 respondents said that 66% of problems assigned to them were solved without involving other people. A corollary is that they would involve others in 34% of the problem cases assigned to them.

Figure 58: Percentage of problems assigned to TS2 participants that are solved without involving others.
In contrast, PG1 TS1 respondents said that only 35% of problems assigned to them were solved without involving others (see Figure 59, LHS). More similarly, PG2 TS1 respondents said that 59% of problems assigned to them were solved without involving other people (see Figure 59, RHS).

**Figure 59: Percentage of problems assigned to TS1 PG1 and PG2 participants that are solved without involving others.**

Respondents identifying themselves as having *novice* or *junior* expertise gave an average response of 37% (see Figure 60, top left) whereas respondents identifying themselves as having *middling*, *senior*, or *expert* expertise gave an average response of 67% (see Figure 60, top right).

**Figure 60: Percentage of problems assigned to ‘novice’ and ‘junior’ participants that are solved without involving others, compared to ‘middling’, ‘senior’, and ‘expert’ participants.**

Researchers Note 5: Since more junior team members rely on their peers more heavily when solving customer problems, perhaps one’s “ability to work as part of a trouble-shooting team” is at least as important as one’s “ability to work alone” identified in question 10.
**Question 13: After what amount of time solving a customer problem would you try to get help from your team-mates or other people?**

TS1 PG1 and PG2 respondents said that they would persist on their own for up to 2 hours, depending on the severity of the case. TS2 PG1 respondents said that they would persist on their own until between 15 minutes and 4 hours, depending on the severity of the case.

Respondents describing their level of expertise as *novice* through to *senior* would wait up to 2 hours, whereas those describing their expertise as *expert* would wait from 2-4 hours to up to 2 days depending on the problem area and impact / severity.

One respondent offered the comment “I tend to keep the case with me if I can”.

**Question 14: What percentage of customer problems assigned to you do you think are brand new, never been seen before?**

Responses to this question are shown in Figure 61. The 18 survey respondents on average thought that about 36% of problems assigned to them were brand new, never been seen before. A corollary is that about 64% of problems assigned to them had been previously seen by themselves or others at HTG.

**Figure 61: Percentage of incoming problems that are brand new – all groups.**
Researchers Note 6: If around 64% of incoming problems are duplicates, perhaps an expert-system style corporate memory could allow 64% of incoming problems to be solved automatically.

On average, TS2 respondents felt that 39% of problems assigned to them were brand new (see Figure 62, LHS), whereas TS1 respondents felt that only 33% of problems assigned to them were brand new (see Figure 62, RHS). A corollary is that about 67% of TS1 problems are duplicates whereas about 61% of TS2 problems are duplicates.

Figure 62: Percentage of incoming problems for TS2 and TS1 groups that are brand new.

Researchers Note 7: The results were over a large spread and may not be statistically significant enough to form a solid conclusion. Perhaps the TS1 survey respondents confused the question, imagining that every new case is a new problem by its very nature?

**Question 15: What percentage of customer problems assigned to you do you think will be seen again yourself or others at HTG?**

On average, survey respondents thought that 76% of customer problems assigned to them would be seen again (see Figure 63).
On average, TS2 respondents thought that 81% of customer problems assigned to them would be seen again, whereas TS1 respondents thought that 70% of their assigned customer problems would be seen again.

Of the TS1 respondents, the PG 1 group thought that 85% of customer problems assigned to them would be seen again (see Figure 64, LHS), whereas PG2 respondents thought that only 58% of their assigned customer problems would be seen again (see Figure 64, RHS).

Researchers Note 8: Compare this with the number of respondents using SolutionDB to store solutions.

Researchers Note 9: Given the higher repetition rate for the PG1 product group compared to PG2, perhaps this is more fertile ground for a FastFIX style solution.
**Question 16:** What percentage of solutions that you create and apply do you think are brand new, never been conceived of before?

On average, TS2 respondents thought that 56% of solutions that they create are brand new, never been conceived of before (see Figure 65, LHS). Compare this with just 19% for TS1 respondents (see Figure 65, RHS).

The corollary is that 44% of solutions created and applied by TS2 respondents are duplicate solutions, whereas 81% of solutions created and applied by TS1 respondents are duplicate solutions.

**Figure 65: Percentage of solutions created by TS2 and TS1 respondents that are brand new.**

On average, respondents identifying themselves as senior or expert in their expertise thought that 58% of solutions that they create are brand new, never been conceived of before (see Figure 66, RHS). Compare this with 21% for respondents identifying themselves as novice or junior or middling in their expertise (see Figure 66, LHS).

The corollary is that 42% of solutions created and applied by respondents identifying themselves as senior or expert were duplicate solutions, whereas 79% of solutions created and applied by respondents identifying themselves as novice or junior or middling in their expertise were duplicate solutions.
Researchers Note 10: Therefore if the process of problem identification and solution matching can be automated, this represents a huge potential area of cost savings for the support centre at both the TS2 and TS1 levels, and for both more and less experienced staff.

**Question 17: When a customer problem is assigned to you, in what percentage of cases would you know the solution straight away, without using other sources of information?**

On average, survey respondents answered that when a customer problem was assigned to them, in 33% of cases they would know the solution straight away, without using other sources of information (see Figure 67). In other words in 67% of cases they would need to refer to other information to solve the case.

**Figure 67: Percentage of cases where survey respondents would know a solution straight away – all groups.**

[Graph showing the percentage of cases where respondents knew the solution straight away]
On average TS2 survey respondents answered that when a customer problem was assigned to them, in 43% of cases they would know the solution straight away, without using other sources of information (see Figure 68, LHS), compared to TS1 respondents whom on average, answered 22% (see Figure 68, RHS).

In other words in 57% of cases TS2 respondents would need to refer to other information to solve the case, and in 78% of cases, TS1 respondents would need to refer to other information to solve the case.

Figure 68: Percentage of cases where TS2 and TS1 respondents would know a solution straight away.

As can be reasonably expected, Figure 69 indicates that more experienced respondents know the solution more readily (45% of cases) than less experienced respondents (23% of cases).

Figure 69: Percentage of cases where less and more experienced respondents would know a solution straight away.

Researchers Note 11: This highlights the importance of accessible sources of “other information” to all groups and most particularly to the TS1 group and to less experienced personnel.
Question 18: If a customer problem comes along, and you don’t know the solution straight away, in what percentage of cases would you first check to see if HTG has experienced that problem before?

Survey respondents indicated on average that when a customer problem comes along, and they didn’t know the solution straight away, in 88% of cases they would first check to see if HTG had experienced that problem before (see Figure 70).

A corollary is that about 12% of cases would be solved from without any reference to the corporate memory.

Figure 70: Percentage of cases where staff check to for previous experience of the problem – all groups.

Question 19: If you do sometimes check to see if HTG has experienced a particular customer problem before, how you actually check that?

Respondents check HTG’s corporate memory using a variety of methods including asking colleagues; using the SolutionDB knowledge base; referring to documentation including product documentation, engineering documentation, T2 documents, release notes, the support matrix, internal service notes, PG2 service alerts, or documents in Docco; checking recent or historical emails; checking their own cases or cases of their peers; searching the engineering database (remedy); checking the clear quest bug tracking system; using instant messenger to contact others; or using Google desktop to search all emails and files on their desktop.
Researchers Note 12: Questions 18, 19, 20, and 21 highlight the dispersed nature of the sources of “other information” required to solve customer problems.

**Q. What type of resources are used to solve problems?**

**Question 20: How often would you rely on the following resources when solving customer problems? Please add and rate any other resources that you rely on when solving customer problems**

A selected list of resources relied upon when solving customer problems was presented to respondents. Respondents were invited to identify how frequently they relied upon each of these resources. A Likert scale from 1 (never) to 5 (always) was used. Responses were ranked using a linearly weighted score as shown in Table 26.

*Table 26: Scoring the most frequently relied upon resources*

<table>
<thead>
<tr>
<th>Response</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>never</td>
<td>1</td>
</tr>
<tr>
<td>rarely</td>
<td>2</td>
</tr>
<tr>
<td>sometimes</td>
<td>3</td>
</tr>
<tr>
<td>often</td>
<td>4</td>
</tr>
<tr>
<td>always</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 27 presents the results in order of most to least frequently relied upon resources.
Table 27: Most to least relied upon information resources – all groups.

<table>
<thead>
<tr>
<th>Resource</th>
<th>Weighted Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>My problem solving skills</td>
<td>84</td>
</tr>
<tr>
<td>My knowledge</td>
<td>83</td>
</tr>
<tr>
<td>SolutionDB entries</td>
<td>75</td>
</tr>
<tr>
<td>Discussion with peers</td>
<td>71</td>
</tr>
<tr>
<td>My experience in the field</td>
<td>62</td>
</tr>
<tr>
<td>Internet searching</td>
<td>59</td>
</tr>
<tr>
<td>Engineering presentation documents</td>
<td>57</td>
</tr>
<tr>
<td>Engineering Technical Manuals</td>
<td>57</td>
</tr>
<tr>
<td>Training course handouts</td>
<td>54</td>
</tr>
<tr>
<td>External Corporate Knowledge Bases e.g. Microsoft, Sun</td>
<td>51</td>
</tr>
<tr>
<td>Customer Manuals</td>
<td>50</td>
</tr>
<tr>
<td>Historical CaseDB cases</td>
<td>43</td>
</tr>
<tr>
<td>Emails</td>
<td>27</td>
</tr>
<tr>
<td>T2 Documentation</td>
<td>21</td>
</tr>
</tbody>
</table>

According to this scoring mechanism, Figure 71 illustrates the top 4 most frequently relied upon resources by survey respondents when solving customer problems:
Figure 71: Top 4 most frequently relied upon resources – all groups.
**Question 21:** Please list in order of most relied upon, to least relied upon, the HTG intranet sites that you visit when solving customer problems

In order of most to least frequently cited, Table 28 lists the HTG intranet sites relied upon by respondents when solving customer problems.

**Table 28: Most frequently identified intranet sites used to solve customer problems – all groups.**

<table>
<thead>
<tr>
<th>Intranet Site</th>
<th>No. of respondents who identified this site</th>
</tr>
</thead>
<tbody>
<tr>
<td>SolutionDB</td>
<td>14</td>
</tr>
<tr>
<td>Docco</td>
<td>12</td>
</tr>
<tr>
<td>Remedy</td>
<td>7</td>
</tr>
<tr>
<td>Aspac</td>
<td>5</td>
</tr>
<tr>
<td>ProductB</td>
<td>4</td>
</tr>
<tr>
<td>SelfServeWeb</td>
<td>4</td>
</tr>
<tr>
<td>SPS</td>
<td>3</td>
</tr>
<tr>
<td>CCA</td>
<td>2</td>
</tr>
<tr>
<td>System Configuration</td>
<td>1</td>
</tr>
<tr>
<td>Product Group 1 (PG1) Technical support</td>
<td>1</td>
</tr>
<tr>
<td>Knowledgelink</td>
<td>1</td>
</tr>
<tr>
<td>Support solutions</td>
<td>1</td>
</tr>
<tr>
<td>Alternate HTG Support website</td>
<td>1</td>
</tr>
<tr>
<td>Lab navigator</td>
<td>1</td>
</tr>
<tr>
<td>Product Group 2 (PG2) support page</td>
<td>1</td>
</tr>
<tr>
<td>CT2 email list</td>
<td>1</td>
</tr>
<tr>
<td>Clearquest</td>
<td>1</td>
</tr>
</tbody>
</table>

Researchers Note 13: Do all respondents within the same group have knowledge of, and access to the same sources of information? Why/why not?
**Question 22: Please list in order of most relied upon, to least relied upon, the external Internet sites that you visit when solving customer problems**

In order of most to least frequently cited, Table 29 lists the extranet sites relied upon by respondents when solving customer problems.

*Table 29: Most frequently identified extranet sites used to solve customer problems – all groups.*

<table>
<thead>
<tr>
<th>Intranet Site</th>
<th>No. of respondents who identified this site</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google.com</td>
<td>12</td>
</tr>
<tr>
<td>Microsoft / microsoft.com / support.microsoft.com</td>
<td>8</td>
</tr>
<tr>
<td>sun.java.com / sun.com / sunsolve</td>
<td>7</td>
</tr>
<tr>
<td>linux / redhat.com</td>
<td>3</td>
</tr>
<tr>
<td>www^196.com</td>
<td>2</td>
</tr>
<tr>
<td>Oracle</td>
<td>2</td>
</tr>
<tr>
<td>unix</td>
<td>2</td>
</tr>
<tr>
<td>Internet RFC sites / <a href="http://www.rfc-editor.org">www.rfc-editor.org</a></td>
<td>2</td>
</tr>
<tr>
<td>xxx.com</td>
<td>1</td>
</tr>
<tr>
<td>yyy.com</td>
<td>1</td>
</tr>
<tr>
<td>HP</td>
<td>1</td>
</tr>
<tr>
<td>HBA vendor sites</td>
<td>1</td>
</tr>
<tr>
<td>zzz.com</td>
<td>1</td>
</tr>
</tbody>
</table>

Researchers Note 14: Despite vendors providing their own solution repositories, Google is still the most cited external webspace to visit when looking for solutions. This highlights the importance of accessibility of available information.

---

^196 The www, xxx, yyy, and zzz websites referred to here are aliases for external websites relied upon by trouble-shooters at HTG.
Question 23: What specific information is most critical to you fulfilling your role at HTG?

Respondents identified the following information as most critical to fulfilling their role at HTG:

- the latest information including up to date problems and trends; what are the most common problems with new code ETAs
- an exact problem description including the customer environment; the error messages and their meanings; symptoms of the problem; logs; and hence a good understanding of the problem.
- command syntax; error codes and their meanings; the order of operations of commands.
- a step by step / point format listing of procedures and tasks
- technical documentation; access to a technical trouble-shooting database
- the ProductB tech 2 document: this is a reverse engineered document on how to do things
- ProductB features and functionality
- engineering based information on how the product actually works, specifications etc
- design specifications, technical manuals, external documentation
- information that describes the system internals - the way the system works
- trouble-shooting manual
- information from engineering and training
- product documentation and knowledge bases
- solutions to identical cases

(Note: duplicate responses have been omitted.)
Question 24: What specific communication is most critical to you fulfilling your role at HTG?

Respondents identified the following communication as most critical to fulfilling their role at HTG:

Information Flows:

- communication from engineering to TS2 and engineering to TS1; good relationship with engineering, information about product updates
- efficient responses from colleagues
- customer communication: being able to tell the customer that I don't know what the fix is but I will do research or make enquiries for the fix
- organisational announcements

Information Content:

- new release ETAs
- known bugs
- severity of the cases and customer impact
- the accuracy of the problem description in CaseDB
- common terminology
- CaseDB case text

Tools Used:

- instant messenger (AIM), email, and telephone

(Note: duplicate responses have been omitted.)
Question 25: What specific meetings are most critical to you fulfilling your role at HTG?

Respondents identified the following meetings as most critical to fulfilling their role at HTG:

- weekly update meetings / huddle with TS2; new product meeting
- team and departmental meetings
- staff meetings to keep abreast of new tools
- meetings with the customer to have them regain confidence
- global team meetings
- meeting with my manager where I can express my needs to fulfill my role

(Note: duplicate responses have been omitted.)

Question 26: What specific processes are most critical to you fulfilling your role at HTG?

Respondents identified the following processes as most critical to fulfilling their role at HTG:

- escalation (5 responses, 1 response as escalation to engineering, 1 response as level 1 escalation process) / turnover
- call handling / trouble-shooting (2 responses)
- service level agreements (SLAs), including engineering SLAs (2 responses)
- CaseDB entry / population / CaseDB being filled out well (2 responses)
- SSE procedures / processes
- knowledge retrieval
- technical documentation distribution
- performance review - it directly tells me which aspects of the job I am not correctly fulfilling

(Note: duplicate responses have been omitted.)

Researchers Note 15: One problem with the existing escalation process is that when you hunt for a solution and don’t find it, you don’t know if its because the solution doesn’t exist yet, or
if it’s just that you haven’t found it. This elevates the level of judgement a trouble-shooter must exercise before escalating the problem. If there was a set list of questions to answer, and solutions to try that became exhausted through the course of the trouble-shooting process, the need for escalation and the decision to escalate would be a lot clearer.

**Question 27: What specific tools are most critical to you fulfilling your role at HTG?**

Respondents identified the following tools as most critical to fulfilling their role at HTG:

- a solution database (11 responses) including SolutionDB (9 responses), and a good knowledge base (database) that provides for quick searches (2 responses)
- CaseDB (case tracking) (7 responses)
- documentation (4 responses), including HTG documents / check sheets and reference pages / product documentation that is available on the web, and documentation for the products and features that we support
- Internet / intranet access (3 responses)
- Remedy (2 responses), including Remedy search
- email (2 responses), including CT2 emails
- an editor (2 responses), including a hex and binary editor
- procomm (2 responses)
- Docco (1 response)
- hardware in the lab for problem recreation
- Google desktop
- service alerts
- trillian (instant messenger)
- putty
- tcpdump
- ethereal
Appendix A – Trouble-shooting Survey

- terminal emulator
- phone

Q. How successful is HTG at solving customer problems?

Question 28: What are the most enabling factors that help you to be effective and efficient in your role at HTG?

Respondents identified the following most enabling factors that help them be effective and efficient in their role at HTG. For easier reading, participant responses have been ordered under the following headings by the author.

Environmental factors (10 responses)

- solid training (technical)
- the HTG intranet
- easy access to good trouble-shooting documentation
- large volumes of information available online
- the public access systems e.g. SolutionDB and documents on the SelfServeWeb
- Google desktop
- tools that work
- the remote dial-in feature
- brown bags - peer to peer forum
- access to email, phone, computers, knowledge base, manuals, and the Internet

Team factors (5 responses)

- very good communication and coordination between team-mates / peers
- communication between other departments including prompt feedback from engineering
- consultation with other parties
enough communication paths
access to other people and their opinions

Personal factors (4 responses)

skill level through training
knowledge of the product and confidence in handling customers
communication skills
experience with the product

Researchers Note 16: Respondents identified factors that they felt were important in enabling their effectiveness and efficiency, not necessarily factors that were presently working well. For example, the comments “tools that work”, and “solid training”, and “large volumes of information available online”.

Question 29: What are the biggest roadblocks that stop you being effective and efficient in your role at HTG?

Respondents identified the following biggest roadblocks that stop them being effective and efficient in their role at HTG. For easier reading, participant responses have been ordered under the following headings by the author.

Training / Knowledge (8 responses)
lack of actual product knowledge and use
training on the latest products
training
lack of training / knowledge
poor training - no engineering training; no proper training aimed at L2 PG1
skillsets, training and basic trouble-shooting skills missing from semi-large parts of the organisation
poor field training
• not able to take any in-depth trouble-shooting class given out by EE or senior T2

Accessible Documentation (6 responses)

• lack of documentation of resources
• lack of technical documentation;
• no book of operations to explain how the product works
• lack of official engineering documentation
• access to in depth product knowledge like the theory of operations etc from eng
• After almost 5 years at HTG, I still can't find the information that I need, either because it is inaccessible or not indexed.

SolutionDB / Knowledge base (3 responses)

• SolutionDB is the most difficult to search knowledge base I have ever used

• SolutionDB
• long search time at the output of SolutionDB;

Time pressures (3 responses)

• time!! not enough time in 8 hours work to complete all the work
• too much administrative work
• no time to properly research the problems

Escalations (2 responses)

• TS1 group disturbs me too frequently; unnecessary calls are escalated
• answering the phone from TS1; this should be handled by the manager to filter and delegate; allowing us to continue working

Customer contact (2 responses)

• handling political issues with customer and HTG
• difficulty in reaching customers
Researchers Note 17: Insufficient training, documentation, and hence knowledge were identified as the biggest roadblocks. A corporate memory managed through a workflow integrated expert system could make documentation more accessible, and could make the need for documentation in certain problem areas more transparent, and it could assist with product training. In theory, at the end of each interaction with an expert system, either the system itself learns something, or the user learns something.

**Question 30a: Do you have access to all the people and team meetings that you require to perform your trouble-shooting role effectively and efficiently at HTG?**

As shown in Figure 72, no significant difference was observed between TS2 respondents compared to TS1 respondents, or between respondents of different expertise.

*Figure 72: Sufficiency of access to people and team meetings – all groups.*

<table>
<thead>
<tr>
<th>Access</th>
<th>Yes</th>
<th>No</th>
<th>Undecided</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>10</td>
<td>2</td>
<td>5</td>
</tr>
</tbody>
</table>

**Question 30b: If NO or UNDECIDED, what additional access to people or meetings would you like, and why?**

**Yes**

- more TS2 interaction in understanding what a particular process / log means / does

**No**

- many of the meetings are help in the US time zone
• access to developers; access to training sessions conducted in other geographies because of the time-zone differences; team meetings

Undecided

• some meetings are scheduled in the early hours due to time-zone issues
• more engineering training required
• during the 'newbie' phase, more access to level 2 staff
• not sure, still new here
• have not been here long enough to know

Question 31a: Have you had all the training that you require to perform your trouble-shooting role effectively and efficiently at HTG?

Responses to this question are shown in Figure 73, Figure 74, and Figure 75.

Figure 73: Sufficiency of access to training – all groups.
Researchers Note 18: The TS2 group were more negative about the sufficiency of training than the TS1 group. As well, more experienced personnel were more negative about the sufficiency of training than less experienced personnel.

Question 31b: If NO or UNDECIDED, what additional training would you like, and why?

Yes

- No comment.

No

- HTG cannot provide training at the technical level required.
• Engineering training on the system internals - without understanding the internals, it’s very hard to solve tricky problems.

• Product training doesn't exist.

• No back-end training which is essential to understanding how the product works as a whole.

• Relevant training for TS2.

• In depth networking protocols; engineering training on new products.

• It could be more efficient - if we had access to the training and informal knowledge transfer of our overseas counterparts i.e. product training.

• PG2 trouble-shooting and PG2 management.

• Basic user training on the products; log reviewing.

• Training on 1. connectivity, 2. devices types; and 3. logs - how to read them: cases are based on them.

**Undecided**

• I have never taken any xxx\(^{197}\), yyy or zzz training before and I don't think there is any available.

• More 'on the job' training is required. This is hard to achieve, as TS1 PG1 is a new team in Sydney.

• Still new here.

Researchers Note 19: Lack of relevant training was strongly identified in the responses. A FastFIX style expert system with corporate memory could help to train trouble-shooters by directing them to relevant sources of information as they solve customer problems. At the end of each case, either the expert system learns something new (for others to share next time), or the expert learns something new (contributing to their training and hence personal development and satisfaction).

\(^{197}\) Aliases have been substituted for the referred products and technologies.
**Question 32a: Do you have access to all the necessary information that you require to perform your trouble-shooting role effectively and efficiently at HTG?**

Responses to this question are shown in Figure 76, Figure 77, and Figure 78.

*Figure 76: Sufficiency of access to information – all groups.*

*Figure 77: Sufficiency of access to information in the TS2 group compared to the TS1 group.*
Researchers Note 20: The TS2 group were more negative about the sufficiency of information than the TS1 group. As well, more experienced personnel were more negative about the sufficiency of information than less experienced personnel.

**Question 32b: If NO or UNDECIDED, what additional information would you like access to, and why?**

**Yes**
- but this information has been generated by TS2!

**No**
- access to in depth details about the product and source code
- all the internal commands are only known to engineering
- book of operations of how the product works
- engineering documentation of HTG products and informal knowledge from engineering
- technical documentation of HTG products is very incomplete... we have to rely on second guessing and reverse engineering the products
Appendix A – Trouble-shooting Survey

Undecided

• it depends on the product/feature - some HTG documentation is not up to date which can cause a lot of confusion
• need access to more detailed information from engineering
• still new here
• I have resources available to ask, but I am often wondering where I can obtain the information myself?

Researchers Note 21: Given the incomplete and sparse nature of technical information, promoting the sharing and reuse of where to find the relevant information could substantially reduce resolution times for repeat problems at both the TS1 and TS2 levels.

**Question 33a: Do you have access to all the necessary software tools, knowledge bases and data bases that you require to perform your trouble-shooting role effectively and efficiently at HTG?**

As shown in Figure 79, no significant difference was observed between TS2 respondents compared to TS1 respondents, or between respondents of different expertise.

*Figure 79: Sufficiency of access to software tools, knowledge bases and databases – all groups.*

![Bar chart showing sufficiency of access](image)
Question 33b: If NO or UNDECIDED, what additional software and tools would you like access to, and why?

Yes

• But not fast or effective all of the time.

No

• not enough lab tools for problem simulation

• customer and TS2 emails should be compiled into a knowledge base

Undecided

• still new here

• not sure if they exist but I would like better tools to remove a lot of the trawling through logs that we currently have to do

• I am sure that there are more tools on the extranet that we just don't know about.

Researchers Note 22: FastFIX could be used to minimise the time trawling through logs.

Question 34a: How effective are you and your peers at resolving customer problems?

Responses to this question are shown in Figure 80.

Figure 80: Customer problem resolution effectiveness – all groups.
Question 34b: Can you suggest any ways to improve the effectiveness and efficiency of your team at resolving customer problems?

Ineffective

• No responses.

Struggling

• No responses.

Middling

• more time in working on problems. Quality rather than quantity.
• all TS2 team members need to know the system internals
• more relevant training; book of operations; access to code
• as it is a brand new team, time will make us all better at problem solving. More mentoring would also help.

Effective

• More transparent error messages should be logged in the log files when a task is forced. Also need to make log files easier to read.
• centralised documentation of solutions and monthly sharing of difficult cases experienced that month
• better access to detailed technical documentation
• access to information and more training; also involving more from the field
• more discussion
• more training
• maybe having different people trained more deeply in difficult areas of our product range
• data organisation could be better
Appendix A – Trouble-shooting Survey

**Very Effective**

- not having to answer the phone from TS1 escalations. This should be filtered and delegated by the manager
- more time to mentor the more junior team members
- engineering telling TS2 how products are meant to work!

Researchers Note 23: A collaborative learning expert system such as that provided by the FastFIX architecture could allow immediate sharing of the high volume, high impact, recent and difficult cases experienced by both trouble-shooters and customers across the globe.

**Question 35a: How effective are you and your peers at solving technical issues?**

Responses to this question are shown in Figure 81.

**Figure 81: Technical Issue resolution effectiveness – all groups.**

![Figure 81](image)

'ineffective': 0
'struggling': 0
'middling': 9
'effective': 7
'very effective': 6

**Question 35b: Can you suggest any ways to improve the effectiveness and efficiency of your team at solving technical issues?**

**Ineffective**

- No responses.
Appendix A – Trouble-shooting Survey

**Struggling**

- No responses.

**Middling**

- more training

**Effective**

- More transparent error messages should be logged in the log files when a task is forced. Also need to make log files easier to read.
- centralised documentation of solutions and monthly sharing of difficult cases experienced that month
- more training

**Very Effective**

- again having specialists in the sub products so you know where the better knowledge is required.
- engineering telling TS2 how products are meant to work!
- have a brown bag / brainstorming issues to share information
- more time to mentor junior members of the team so that they can develop their technical skills

Researchers Note 24: For questions 34a and 35a, 13 of 17 respondents answered either ‘effective’ or ‘very effective’. However, in comparing responses in this question with the previous question, respondents have greater confidence in resolving technical issues, than in actually solving customer problems.

Researchers Note 25: A collaboratively created corporate memory in the form of an expert system would facilitate a centralised sharing of solutions far beyond the limited scope of SolutionDB. It would allow solutions on the public Internet, within Docco and the corporate intranet, within vendor knowledge bases on the web, and in SolutionDB and Remedy to be readily indexed and made available within the scope of any given problem on hand.
Question 36a: How effective is the current software suite at HTG at facilitating an efficient an effective flow of work through the support centre?

Responses to this question are shown in Figure 82 and Figure 83.

Figure 82: Effectiveness of the current software suite – all responses.

![Figure 82: Effectiveness of the current software suite – all responses.]

*ineffective*: 0  
*struggling*: 3  
middling*: 7  
effective*: 5  
*very effective*: 1

Figure 83: Effectiveness of the current software suite – TS2 compared to TS1 responses.

![Figure 83: Effectiveness of the current software suite – TS2 compared to TS1 responses.]

*ineffective*: 0  
*struggling*: 3  
middling*: 3  
effective*: 2  
*very effective*: 1  
*ineffective*: 0  
*struggling*: 0  
middling*: 4  
effective*: 3  
*very effective*: 0
Appendix A – Trouble-shooting Survey

Question 36b: Can you suggest any ways to improve the effectiveness and efficiency of workflow through the support centre by changes to the software suite?

Ineffective

• No responses.

Struggling

• abandon CaseDB and SolutionDB; integrate better - better ability to search; central place to go where you can do everything that the CaseDB / SolutionDB / remedy / interoperability matrices are trying to do - no centralised standardised mechanisms

• replace SolutionDB with something that works and make CaseDB searchable

• have 1 system

Middling

• the software suite is inefficient and needs better quality control

• SolutionDB has many junk items, CaseDB needs to be integrated into SolutionDB

• there are too many places that we have to search in

• reduce duplication of email / case text - one click / one op does it all! still not available

• stop using multiple disjointed case logging tools - CaseDB, online problem tracking (opt), clearquest etc

• software for call tracking is good but we are suffering from slow processing time due to the network connection speed to the server which is in the US

Effective

• have an introduction / training day to introduce the tools to new staff

Very Effective

• No comments.

Researchers Note 26: The net answer here was that the present software suite is middling. TS2 respondents were more negative about the current software suite than TS1 respondents. FastFIX is designed to provide a lightweight brokering architecture that can neatly fit between
web-enabled case tracking, knowledge base, or documentation systems, indexing problems to their solutions and thereby recording and reusing expert problem determination and search knowledge.

**Question 37: If you had unlimited budget and ICT support to build one or more software tools to help you do your job better, what would you wish for?**

*Integrated and Searchable Knowledge base, Case Tracking, and Document management (12 responses)*

- better knowledge base, at this time SolutionDB has a poor search engine
- make SolutionDB better
- more intelligent searching for SolutionDB
- **One knowledge base.** Just one search engine to search SolutionDB, SelfServeWeb, Docc, and previous CaseDB cases.
- **an integrated tool combining call tracking, knowledge base document repository etc**
- Google / fast / noise reduced tool that gets fewer hits but are precise
- a search algorithm as strong as Googles; **one DB that contains everything:** cases; manuals; solutions etc
- artificial intelligence (AI) - although unlimited budget since the 50s has still failed to deliver
- CaseDB, SolutionDB, and time reporting tools and remedy should be integrated into **one solution**
- **a single case management tool** that everyone uses, a webcrawler and index / search for the intranet
- **1 system**
- much better CaseDB where you can do word searches for errors on a local replica and get % hit out
Automatic Log file analysis (2 responses)

- a log scanner that matches errors with a database of solutions
- automatic log file analysis

Other comments

- every type of product available for hands-on testing
- ftp straight to homer not down to the desktop then back up to homer; time downloading SAlerts or patches and logs
- I wish that the PG1 code itself worked correctly!
- documentation, people, time
- good training

Question 38: If you had unlimited budget for additional resources to help you do your job better, what would you wish for?

Responses have been split into the following subheadings for easier identification.

More People (12 responses)

- one or two more TS1 staff to mentor us.
- tech writers who would write a book of operations and uncovered series for PG1
- quality (TS2 standard) new hires
- more staff
- an additional engineer
- appropriate staff levels
- more engineers to help reduce the workload so that we could improve our work quality and be able to spend more time on research and problem recreation
- more junior staff to mentor to allow more senior staff to write documentation
- people - to provide better feedback to developers
• expand the TS1 organisation and foster them up to a higher level
• a couple more well-trained PG1 personnel
• a person to help me out

Training (5 responses)
• more time in training; especially mentoring
• on site eng training
• training levels in all departments; product specific e.g. redhat linux training
• more training
• more training

Equipment (3 responses)
• more WAN bandwidth
• faster equipment
• an extra pc / screen for working the case and a dial-in; a larger desk area for looking through docs;

Log scanning
• more time to review logs or an application that could do a base scan of all the logs and locate errors and maybe match with a knowledge base for possible resolutions

Relationships
• closer relationship with engineering

Knowledge base
• Google / fast / noise reduced tool that gets fewer hits but are precise

Researchers Note 27: The common theme here is a desire for more staff and a better process for developing them. The FastFIX architecture provides a forum which allows more junior staff to be developed and the expertise of more senior staff to be leveraged.
Assessing CaseDB’s Fitness for Purpose

Question 39a: Please list the tasks that you perform in CaseDB from day to day, and insert an estimate of the average daily time that you spend performing each of these tasks.

On average, respondents spend 3 hours and 9 mins each day using CaseDB. The tasks that respondents perform in CaseDB from day to day include:

- case admin - monitoring queues, accepting cases, creating cases, assigning cases, case dispatch - closure or assignment back to the field, CAC / turnover (15% - 29 mins per day)
- searching for customer info e.g. locating modem and serial numbers (5% - 9 mins per day)
- reviewing the case and case history to determine the current status (13% - 25 mins per day)
- analysing and entering technical data / research text / escalation results (19% - 35 mins per day)
- entering case updates - customer correspondence + environment screen (45% - 1 hour and 26 mins per day)
- searching old tickets (3% - 5 mins per day)

Question 39b: Total Time spent each day on CaseDB (hours)

Responses: 17, Range: 1.5 to 8.0, Average: 3.9, Median(s): 4 (see Figure 84).

Figure 84: Total Time spent (hours) each day on CaseDB – all groups
Question 40: What would you say is the purpose of CaseDB in the overall HTG workflow?

- case tracking and billing application; customer database
- great call tracking system, different level of groups (ie management, engineering) able to extract data from the database for reporting purposes
- case history
- record keeping / statement of proof (accountability)
- case organisation and tracking
- tracking customer information and cases
- auditing and communication
- case (problem) management and tracking
- keep track of case workflow including details of the latest updates
- to collect a case history so that anyone can pick up a case and work on it
- to track the case and update all the relevant parties in a timely manner
- central place for case specific exchange of information
- to track case issues from conception to resolution
- problem reporting and progress reporting
- keep everybody up to date with case progress
- support requirements; tech support and field operations
- an interface for customers to see the current status, and information transfer medium between the tech staff, a timeline for root cause analysis and [please explains]!

Researchers Note 28: CaseDB is basically a case-tracking database. It stores incoming problem tickets.

Question 41: How would you rate CaseDB overall in terms of its fitness for purpose?

The responses to this question are shown in Figure 85.
Researchers Note 29: The overall assessment of CaseDB’s fitness for purpose was ‘reasonable’ to ‘good’.

**Question 42: Please describe the top 3 things you like about CaseDB (pros).**

- customer information / case issue-problem; current work status - history; link to SolutionDB
- customer viewable text; customer able to update the case
- easy to use, record keeper, searchable
- ability to track cases
- lots of information available
- the layout, the searching facilities, aim keys
- contiguous case text; hardware list; previous cases, by serial number
- easy use; flexible
- case history, equipment information
- none
- information about the case is centralised at a single point
Appendix A – Trouble-shooting Survey

- nothing
- can obtain a history of the event
- customer sees updates in CaseDB instead of calling / disturbing me; the information can help with future problems
- a central place to put updates; management loves it!
- reasonably stable, easy to use, everyone has access to it

**Question 43: Please describe the top 3 things you dislike about CaseDB (cons).**

- having to re-enter the same text; speed / time opening up windows e.g. environment screen
- slow; no flag shown if environment screen has been populated or not
- slow
- duplication (avoidable!), slow
- too many things to input; slow; some inputs are irrelevant
- slow, slow, slow!
- the speed, incorrect information (modem numbers), having to duplicate information, the number of windows that need to be opened
- slowness; environment screen requires filling out each case; only as useful as the case text
- no easy search through other cases for keywords
- locating customer information - it is spread everywhere
- no word search capability; slow and cumbersome; irrelevant data in there which is often inaccurate; no trends; sometimes can't close case or yank
- it is slow; people fill it with irrelevant cases; people sometimes fail to update it
- speed, ability to search
- there is virtually no technical data
Appendix A – Trouble-shooting Survey

• too slow - but has been improved
• slow; not very flexible – the custom macros demonstrate that; not very user friendly
• slow over WAN, upgrades for CaseDB always occur during the Sydney shift, poor support when something doesn't work

Researchers Note 30: The FastFIX system would ensure that irrelevant data was never recorded or requested – this would effectively replace the present CaseDB environment screen. The FastFIX system would replace the need for custom macros. The FastFIX system would encourage technical data to be captured and questions asked as is relevant for the problem at hand.

Question 44: How long on average each day would you spend looking for CaseDB cases to compare the current case against?

• never (7 responses) - there is no good facility for it
• not at all unless it is obvious that another case exists
• depends. I only go back to old cases if I have a record of them.
• not long - 5 mins
• 30 mins
• 30 mins or 1 hour
• 1 hour

Question 45: There are many vendors offering solutions for the tracking customer problems at the help-desk / call-centre / support-centre. Can you name any?

• AR Remedy / royal blue / iTAM; heat (not HTGS one), peregrine, CaseDB, PAL, sebel, help desk, onyx, ibm, hitachi, sun, CPMA (mainframe - IBM), CA unicentre
Question 46: Are there any vendor solutions for tracking customer problems that you would recommend? What is your experience with them? Why would you recommend them?

- remedy / less noise
- I have always used in house solution - I find every company has different needs - so an in house solution is the best choice.
- none specifically; all much the same
- IBM have a good one
- Google has a pretty good search engine

Researchers Note 31: A FastFIX solution could provide a map that charts the course between the problem a customer has, and the information that the customer needs, with a better and more targeted closed-loop feedback data matching system than that provided by Google.

Assessing SolutionDB’s Fitness for Purpose

Question 47a: Please list the tasks that you perform in SolutionDB from day to day, and insert an estimate of the average daily time that you spend performing each of these tasks.

On average, respondents spend 77 mins each day using SolutionDB.

The daily tasks that respondents perform in SolutionDB include:

- searching for existing solutions (92% - 71 mins per day)
- creating new solutions (6% - 5 mins per day)
- linking existing solutions to the CaseDB case on hand (2% - 1 min per day)
- approving SolutionDB solutions (0% - 0 min per day)

Question 47b: Total Time spent each day on SolutionDB (hours)

Responses: 17, Range: 0.1 to 2.5, Average: 1.4, Median(s): 1.5 (see Figure 86).
Figure 86: Total Time (hours) spent each day on SolutionDB – all groups

Question 48: What would you say is the purpose of SolutionDB in the overall HTG workflow?

- knowledge base of solutions for errors that have been found
- with the new SelfServeWeb, SolutionDB and integration, it is supposed to be easier for the customer to open a case in the event that the customer cannot find a fix to their problem. However, based on the customers survey report, feedback suggests that SolutionDB needs much improvement.
- to make our job easier by reducing research time
- clear queues faster by providing clear solutions
- global searching of knowledge; less time in solving problems and looking for solutions
- to provide access to solutions for historical cases. It does not do this well.
- a problem / solution knowledge base
  - exact error library -> link error to the solution
- track existing solutions especially for level 1 support
- to collate solutions to similar problems
- keep track of known problems to reduce duplication of effort
Appendix A – Trouble-shooting Survey

- a failed searchable database of solutions - *searching is too complex, and there is too much noise in the output*

- to create a prepared solution of your issue in order to help others

- to search for existing solutions to the same or similar problems

- searching for a matching solution in a timely manner

- minimise the time for solving a problem

- knowledge sharing for break/fix issues

Researchers Note 32: SolutionDB is basically a solution storage database. It stores solutions to known problems.

Researchers Note 33: Neither CaseDB nor SolutionDB record either the problem determination knowledge that a trouble-shooter uses when determining the ‘Class of Problem’ on hand, or the search knowledge that a trouble-shooter applies when hunting for the solution to a given class of problem.

**Question 49: How would you rate SolutionDB overall in terms of its fitness for purpose?**

Responses to this question are shown in Figure 87.

*Figure 87: SolutionDB’s fitness for purpose – all groups*
Researchers Note 34: The overall assessment of SolutionDB’s fitness for purpose was ‘weak’ to ‘reasonable’.

**Question 50: Please describe the top 3 things you like about SolutionDB (pros).**

- auto entry if opened from the case; query options
- folder for quick access to known issues; search constraint filter
- many search criteria: 1. easy search; 2. organised by product segment; 3. concise
- a lot of data. you can drill it down to specifics
- well documented answers
- detail of information (release levels are specified), ability to link to CaseDB
- provides perfectly tested fixes, ranked according to relevance, easy to add multiple search strings
- easy search
- restricting to specific products via query filtering; different levels of viewing; web based
- matching statements facility
- none
- solutions are prepared, the search is reasonably fast
- it’s ok for well known problems, there’s virtually nothing I like about it!
- easy to use; fast; can update the solution in real time
- sometimes I get what I am looking for, other than that I do not like it
- works in a web browser, no installation needed, fairly stable

**Question 51: Please describe the top 3 things you dislike about SolutionDB (cons).**

- list of solutions returned; viewing the returned solutions that matched with the input
Appendix A – Trouble-shooting Survey

- poor search engine; sometimes search result shows solutions that are not finished
- slow: 1. *too much noise*; 2. *common terminology not used on all solutions*
- difficult to search; results are sometimes irrelevant; solutions are sometimes wrong
- slow; search is very poor
- hard to search, hard to bring back previous solutions, *uncertainty of the existence of an answer*
- continually have to re-login; if no fix exists it provides 0% help, can be slow / offline
- articles might be out of date; speed - slow
- too general; hard to find solutions; no Google search
- searching is too complex; solution matching is poor; *too many false positives*
- some of the items are junk; would like it to be faster
- *there aren't many usable solutions; there are many junk solutions*; it has a weak search engine
- the search is no good; requires training to search; slow an takes time to create a search
- searches are difficult and unreliable, slow over WAN, non-intuitive interface

**Question 52: How long on average each day would you spend looking for SolutionDB solutions (hours)?**

Responses: 17, Range: 0.1 to 2.5, Average: 1.1, Median(s): 1 (see Figure 88).
Question 53: When you’ve got a problem on hand, and it appears to be a variation on a problem that has been seen before, which of the following options do you prefer?

Responses to this question are shown in Figure 89.

Note: 3 of the 4 respondents answering ‘neither of these’ commented that they were new to HTG and therefore had no experience with either creating or editing SolutionDB solutions.

One respondent said the best strategy was to contact the original SolutionDB author and have them edit their own solution.
Researchers Note 35: It appears that a good number of users would rather not record solutions in SolutionDB at all, or create possible duplicate solutions, rather than go through the pain of editing an existing solution.

**Question 54a: How comfortable are you with editing a SolutionDB solution that someone else has created?**

Responses to this question are shown in Figure 90.

*Figure 90: Comfort with editing SolutionDB solutions – all groups*

Researchers Note 36: On balance, users are uncomfortable to neutral with editing a SolutionDB solution that someone else has created.
Question 54b: If you are uncomfortable with editing a SolutionDB solution that someone else has created, please explain why:

- my knowledge is too limited to write a solution just now (4 responses)
- unless they are obviously wrong I would hesitate in case I erred.
- better to create a new SolutionDB; can clutter existing solution with too much data
- I normally do not edit, I normally get the author to edit it
- think we should notify and request the author / SolutionDB owner to modify according to the latest fix and possibly give the CaseDB case # as a proof.
- need to contact the original author, check the environments are the same, make sure my findings don’t clash with the original findings, perform extra investigation

Researchers Note 37: The last response seems to make the most sense, assuming that the author is still in that role at the company. Perhaps this approach should be encouraged more formally?

Question 55: Before you create a new SolutionDB solution, how often would you check that a similar solution doesn’t already exist in SolutionDB?

Responses to this question are shown in Figure 91.
Researchers Note 38: While the great majority of respondents claim to ‘always’ check that their new solution will not duplicate an existing SolutionDB solution, the ‘sometimes’, ‘often’ and ‘usually’ responses indicate that there is some weakness in the workflow. Users can search endlessly and not know if they just haven’t yet found the right solution, or if in fact the right solution just doesn’t exist. That may explain some of the frustration experienced when searching to avoid duplicate entries. This would be minimised with the FastFIX architecture since for a given problem class, the conclusions will either explicitly be there, or they won’t be there at all.

**Question 56: Before you create a new SolutionDB solution, how much time (minutes) on average would you spend checking to see if a similar solution doesn’t already exist in SolutionDB?**

Responses: 11, Range: 0.0 to 60.0, Average: 27.5, Median(s): 10 (see Figure 92).
Figure 92: Time (mins) spent checking SolutionDB prior to solution creation – all groups

Question 57: How long (minutes) on average each day would you spend editing SolutionDB solutions?

Responses: 12, Range: 0.0 to 5.0, Average: 0.8, Median(s): 0, 0 (see Figure 93).

Figure 93: Time (mins) spent editing SolutionDB solutions – all groups
Appendix A – Trouble-shooting Survey

**Question 58: How long (minutes) on average each day would you spend creating SolutionDB solutions?**

Responses: 13, Range: 0.0 to 5.0, Average: 1.5, Median(s): 0 (see Figure 94).

**Figure 94: Time (mins) spent creating SolutionDB solutions – all groups.**

Respondent’s Comment: I don't create SolutionDB solutions everyday. Maybe it would take around 2 hours if I really had to create one.

**Question 59: There are many vendors offering solutions for knowledge management at the help-desk / call-centre / support-centre. Can you name any?**

- SolutionDB; microsoft technet; sun sunsolve
- remedy, professor plum, remedy/heat; royal blue
- sunsolve and the knowledge base on the www.concord.com site
- CA unicentre
- SolutionDB
- SolutionDB, CaseDB, peregrine, remedy
Question 60: Are there any vendor solutions for knowledge management that you would recommend? What is your experience with them? Why would you recommend them?

- As with case logging, my experience is with in house systems that do both in house. It's the only way to get a system that suits your every need.
- remedy
- Google has a search engine that works
- Google desktop search

Question 61: If you create a SolutionDB solution, are you confident that you will be able to find it again?

Responses to this question are shown in Figure 95.

*Figure 95: Ability to find one’s own SolutionDB solutions – all groups*

Question 62a: How would you rate the SolutionDB search interface?

Responses to this question are shown in Figure 96.
Researchers Note 39: The overall rating for the SolutionDB search interface was ‘weak’ to ‘reasonable’.

**Question 62b: What would you recommend for improving the search mechanism in SolutionDB?**

- more like a web search engine like Google; highlight the match in the supplied solutions
- weak search engine; very slow due to the connection to the solution database in the US. If we could improve the algorithm in the search engine it would help a lot
- common terminology; easy filters
- Google style
- simplify the interface, make searching by 1/2 keywords work
- make it easier to know the standard formatting e.g. do I write: os:windows; operating system: windows; or just windows
- integrate with product manuals and / or engineering resources
- Google search
- replace it with something else
- need to make use of a full text search instead of a keyword search
- the existing mechanism is not good; I like something like Google
- a second Google-style interface just for searching

Researchers Note 40: Attractive features of the Google search engine are lightspeed response due to caching and indexing, and closed-loop feedback that promotes frequently clicked on and referenced pages to the top of the search list for a set of given search criteria. One of the downsides of the Google caching mechanism is that updates to the search engine are done over the matter of days rather than seconds. As well, feedback is based on statistical click-through rates which may or may not correspond to what the searcher was actually hunting for. FastFIX proposes instantaneous closed-loop updates. As well, solutions are expertly linked to the user’s current context.

**Question 63: Please describe how you go about trying to find solutions in SolutionDB**

- If I use the SolutionDB button in CaseDB then facts are added to the SolutionDB search from CaseDB’s environment screen. After that I add the error if available on what I am searching for, then run the query. Often I will filter with the saved query constraints to reduce the solutions returned, then I’ll review the returned outputs. Sometimes I will also look for the matching statements that could represent my question or goal.

- use search constraints for the product that I am working on; enter as much details / facts and symptoms as possible including using the "reuse statements" function

- search using symptom / goals statements

- select the product line; type the version; cut and paste email error message; search

- input the symptom or the error message and the facts i.e. the environment

- I input the search term into SolutionDB and then try and narrow the results down where possible.

- just keep trying! enter a search string and then utilise the 'matching statements' feature.
• type in error message and hit enter
• keyword search or context search
• enter the products relevant data and pray!
• matching statements facility; use Google search to find SolutionDB ID then put into SolutionDB
• put log entries in search field; how do I?; put product information in
• just cut and paste error messages or symptoms and articulate with specific version / product
• it's very simple, you either find a solution in 15 mins or you don't!
• open SolutionDB; search for the words in SolutionDB using (…); SolutionDB only shows the first 20 hits!; I don't want to spend more time doing SolutionDB searching; no training should be required
• using one keyword per line, not normally more than 2-3 keywords in total. Trying different keywords until successful e.g. powerpath, windows 2000

Researchers Note 41: FastFIX proposes to better automate the searching by populating the search criteria more fully in the case of open searches, and by otherwise linking to explicit SolutionDB solutions.

Question 64a: When you use a particular SolutionDB solution to solve a CaseDB case, how often would you link the CaseDB case to the SolutionDB solution?

Responses to this question are shown in Figure 97.
Researchers Note 42: Respondents ‘sometimes’ link a given CaseDB case to the SolutionDB solution that solved it. These results would indicate that the repetition rate of problems documented in SolutionDB is possibly much higher than SolutionDB may indicate. This further encourages the use of a FastFIX style collaborative corporate memory.

**Question 64b: If never or rarely, please explain why?**

- SolutionDB is not user friendly; CaseDB is not designed to work with SolutionDB; extra work; both too slow
- no always a direct match, maybe only a hint of what the problem is
- time is of the essense
- just forget to link (2 responses)
- only a recently live team - PG1 L1

Researchers Note 43: Workflow solutions need to ‘pave the path of least resistance’ – not expect people to jump out of their box!
**Question 64c: How and where do you actually create the link?**

- on the link icon in the toolbar when the solution is opened, but have also used the link option in the return window - solution window / frame
- click the link button on SolutionDB
- CaseDB -> SolutionDB link button
- on the SolutionDB web page click on the chain icon
- via the CaseDB link button
- from within SolutionDB
- in the CaseDB case, use the SolutionDB button
- the SolutionDB button in CaseDB
- refer to it in the case text
- use the SolutionDB button in the CaseDB GUI to insert the relevant SolutionDB number
- using the chain icon in SolutionDB
- no idea

Researchers Note 44: This link would be automatically created by a FastFIX-style system.

**Question 65: Of the solutions that you create in SolutionDB, what percentage of solutions do you think will be applied again by yourself or others at HTG?**

Responses: 14, Range: 10.0 to 100.0, Average: 61.1, Median(s): 50, 70 (see Figure 98).

Respondents note: The solution reuse rate will depend on whether they can find the solution again!
Researchers Note 45: These results suggest only a moderate reuse rate for SolutionDB solutions. The FastFIX solution ensures that people don’t waste time refining non-repeat solutions. With FastFIX, the most frequent classes of case wear the smoothest path through the solution search space.

*Question 66: What percentage of solutions that you create and apply would you record in the SolutionDB knowledge base?*

Responses: 15, Range: 0.0 to 100.0, Average: 52.0, Median(s): 70 (see Figure 99).

Researchers Note 46: Responses to question 47a and b would indicate that this is an overly optimistic response.
Question 67: What would encourage you to use SolutionDB more in your daily work?

- **better returned solutions**, faster refresh rate
- better search engine algorithm
- faster searches; better search mechanism
- **less noise**: common terminology
- better search; **solutions need to be checked to make sure that they are valid**
- if it was faster and gave better results. Google consistently returns the correct result to me in the first page from 8 billion on pages - why can't SolutionDB?
- **If I actually knew, for sure, that a solution existed**
- more detail for every day problems of a non-specific nature
- **better definition of problems and solutions**
- more time
- if it was replaced with something that works
- time
- no time to use it in the current situation
- I am using it everyday already
- quicker, effective Google-like search; able to store word and pdf files
- **some type of recognition for a. creating solutions and b. use of those solutions by peers and customers**

Researchers Note 47: The prototype version of FastFIX presently being trailed already allows users to upload Word and PDF files. The FastFIX solution recognises valuable contributors to the knowledge for example as Gold, Silver and Bronze contributors.
Survey Part B - First Impressions of FastFIX

Question 71: What do you think of the FastFIX model presented?

- Quite cool. A different solution search approach.
- Good to allow more natural ways to discover existing solutions using existing tools.
- Has potential. Would like to see a working production environment scenario.
- I think it is great because sometimes I spend a lot of time just trying to search SolutionDB or the documentation. I am sure it will help me to do that.
- It looks a little like the help option in the microsoft help asking questions and then giving possible solutions with a [did this fix it or not] question. The presented model looks OK as it learns the correct answer.
- This will work with OSAPT / infrastructure team due to the nature of the problem.
- Great tool. Should benefit us here.
- The algorithm is good / workable. Based on the wiki that I have used at USyd, I think it is a great way to implement knowledge management.
- Good idea for generating initial and valid questions to drill down to the main problem.
- It looks quite impressive. I am a fan of expert systems and look forward to having access to FastFIX.
- Paving the path of least resistance is a good idea. Using a 3rd system would not work. A strict parameterisation of problem attributes would help with searching. Linking to the documentation is a great idea.
- The model looks functional and the concept seems to be useful, at least on a theoretical level.
- Interesting. Very dependent on people and their discipline to create and fine-tune the model/rules.
- It appears to address a hole in the system and is much needed.
- Good Idea.
• Interesting. I hope it will work!
• Quiet good.
• The idea is quite good. A lot of effort and development will be required.
• The model is good. It's a good idea to integrate the call management with the knowledge base. I see it working when the attributes associated with the case are measurable. A common terminology will be needed – that can make things complicated. I find that it is easy to do a Google search manually with either my own or someone else’s description of the problem. I think FastFIX will be useful for CSI or level 0 when the volume of calls are very high. TS2 could contribute to the database.
• Looks interesting - would need to have a play with it. The idea seems good.

**Question 72: Is there anything that you didn’t understand?**

**Yes (3 responses)**

• Not sure of integration methods with existing tools.
• I briefly understood what it is. But as I've only just seen it and haven't yet used it, I don't understand how it works for me.
• Exactly how would FastFIX sit between SolutionDB and CaseDB?

**No (16 responses)**

• No comments.

**Undecided (1 response)**

• I guess so. Will be more clear when I start using it.

**Question 73: Do you have any criticisms of, or concerns with, the architecture presented?**

**Yes (6 responses)**

• Do we need to build another knowledge base?
• Not sure how FastFIX would integrate with SolutionDB e.g. rejected solution -> info added to SolutionDB? Build up of FastFIX -> slow down of new SolutionDB solutions?

• TS2 level not sure, perhaps at level 1 space this could work.

• This might be difficult to implement in our group (intelligence supervision - PG2) because of the difficult solutions and symptoms to problems.

• It still relies quite heavily on busy SAC staff to populate the rule base.

• Is very reliant on human input. The time to enter the relevant rules may be skipped due to time pressures.

**No (13 responses)**

• the architecture makes perfect sense.

• the only problem for PG1 would be the time spent entering data.

• It seems to be easily automated. Would be good to get it into HEAT.

**Undecided (1 response)**

• I need greater familiarity and time to know the answer

**Question 74: What improvements can you conceive of? What refinements would you suggest?**

**Yes (6 responses)**

• There should be a descent way to integrate the tool into existing tools without any pain.

• All queries should link to the manual / guide for that product and / or feature. Searching should be allowed in natural language and parameterised.

• Integration with all HTG knowledge repositories, not just CaseDB, SolutionDB, but things like ProductB website, Tech2 guide, Docco, and Remedy. Also, addition of 'how to' type topics, with plain english searches.
Appendix A – Trouble-shooting Survey

- Maybe force the user to input data if the current entries are vague. Also, force them to acknowledge whether it solved the problem.
- A process by which all the entered data is reviewed and approved / rejected by a senior trusted person.
- The tool must not have its own user authentication system - we have enough usernames and passwords already.

None (9 responses)
- none for the presented model. It would have to be refined as it gets used.

Undecided (4 responses)
- Too early to comment.
- Not at this stage - I need to see the real world usage before I can judge it.
- Bit hard to say at this stage.
- It is hard to give any comment at this early stage. We have to wait until the database gets populated and cases start getting refined in the system.

**Question 75: Do you think that the FastFIX approach could help you in your job role?**

Responses to this question are shown in Figure 100.

**Figure 100: Will FastFIX help – all groups?**

![Graph showing responses to Question 75]

'yes': 13
'no': 0
'undecided': 7
Appendix A – Trouble-shooting Survey

Question 76: Why / why not?

Yes (13 responses)

- If it allows me to find the data I need then I won't need to keep my own database anymore!
- If integrated, it would save time used in searching multiple databases.
- My job is to encourage more building up of knowledge - anything to make that easier is good.
- If everyone can start incrementing at least one solution per week, I think this tool would be very useful.
- Because having answers presented before I ask a question will save time.
- If the algorithm works in real life, which it does, it can work anywhere.
- Improved time to resolution
- If it reduces the time to find a solution then it will be useful.
- By linking to SolutionDB / files / documentation it would speed up the search process.
- Wont miss any vital information / data gathering
- It will help me reduce searching time - we have a lot of information and documents in various locations so it's hard to find out exactly what I want in the moment. It will be good to share the information across the team.
- 1. Knowledge transfer between members of a group. 2. knowledge transfer downward - ie from L2 to L1 and to LSC

No (0 responses)

- (There were no negative responses)

Undecided (7 responses)

- Not at a TS2 level, good for level 1 space.
- In PG2 support, almost each case is different. There is not much repetition.
Appendix A – Trouble-shooting Survey

- I do not know how effective the system will be in handling complicated low volume, high impact time consuming problems which require higher levels of technical expertise.

- As with controlCentre the error may be resolved by a number of different solutions. A number of different causes could create the error given.

- Problems that we encounter are rarely granular.

- The product that I am currently supporting has no straight forward answer. It is interdependent on other components.

Researchers Note 48: Applying FastFIX across the global support organisation will minimise the incidence of duplicate problems floating through to TS2. The granularity of problems is already demonstrated by the manner in which cases are encoded through the environment screen in CaseDB, and through the facts data in SolutionDB. We aren’t attempting to codify cases that aren’t able to be codified. The system is designed to smooth the path of handling high volume repeat problems. In the pathology domain, only 1 minute per case is required to populate the rule base with solution links relevant to the problem on hand, irrespective of the size of the knowledge base.

**Question 77: Do you think that HTG should invest in building the FastFIX tool?**

Responses to this question are shown in Figure 101.

**Figure 101: Should HTG invest in FastFIX – all groups?**
Question 78: Why / why not?

Yes (13 responses)

- It has the potential to be very useful.
- Time saver, uniformity etc.
- If this project will help make our jobs easier, I believe HTG will support it.
- Because having answers presented before I ask a question will save time. Less time per case -> more cases per day -> more happy customers -> more sales.
- Would be great in HEAT - would decrease resolution time. This would increase time for solutions that aren't in the system.
- Customer satisfaction.
- There are currently too many places to search for information.
- Both CaseDB and SolutionDB are not enough and this might be the missing link if successful
- I think it will help me to reduce our total workload, especially for the simpler issues or questions.
- I think that a point format / step through type model would help with some of the basic solutions. Maybe it could even be used by the customer on the SelfServeWeb.
- To see if it works - we need to build a system and build a knowledgebase before testing can occur.
- I would like to see big improvements on the SolutionDB search engine and the length of time required to solve a problem.

No (0 responses)

- (There were no negative responses)

Undecided (7 responses)

- I know nothing of this area - if FastFIX is the only solution on the horizon then yes, otherwise consider other tools also.
• Not yet proven in the field

• Let HTG decide!

• I would like to see it working. The idea is good but I would like to see how effective the implementation will be.

• Not sure if it would help with PG1 but may be of use for ProductA hardware problems.

Researchers Note 49: A properly resourced trial with relevant IS access to existing SolutionDB and CaseDB infrastructure, databases and code is the only sure way to see if this idea is as good as it sounds. If HTG doesn’t run with it, someone else will!

**Question 79: Do the ideas presented remind you of anything? Have you seen something like this before?**

**Yes (13 responses)**

**Microsoft (4 responses)**

• Yes - like help in Microsoft. A question followed by suggestions, followed in turn by a [did this fix it] question, and if not, an alternate solution.

• Microsoft self trouble-shooting technique.

• just microsoft trouble-shooting

• Yes. Microsoft's web help asks relevant questions and asks whether it solved your problem.

**SolutionDB (2 responses)**

• I think this idea is very similar with SolutionDB. Just maybe more robust.

• SolutionDB.

**Remedy (2 responses)**

• Yes - PRDF Plum, Remedy / Heat.
• A very old version of remedy. Classification of problems was very strict to build up rules to be able to search on.

Other

• I have seen something like this in my previous company, but it's not exactly the same.
• My uni days.
• Seems to be just applying logical trouble-shooting.

No (7 responses)

• No comments.

Question 80: What web forums / chat rooms / blog sites / discussion boards on the wider Internet or on the HTG intranet would you personally recommend as being effective sites for collaboratively exchanging ideas?

• There is a meeting room, but I have never used it yet.
• email groups
• OE information exchange.
• chat rooms, email
• I sometimes use computing.net
• wikipedia.org, linuxquestions.org, arstechnia.com
• Wider Internet: vendor (HP etc) websites.
• Mailing lists - these tend to have a higher quality of information than the other more interactive mediums.
• PG1 'CT2' email list.
• webex, SolutionDB, instant messenger and daily email is what we use to exchange ideas.
• Wiki where everyone contributes, email lists like symcli expert group, PG2 information exchange etc.
Question 81: Do you have any further questions or concerns?

Yes (1 response)

- Is it possible to populate a system like this with known solutions? Have this done by staff other than those that are trying to take cases at the same time.

No (19 responses)

- No comments.

Researchers Note 50: Yes. It would be possible to do datamining on CaseDB and SolutionDB to populate FastFIX. We would need IS to provide low level password read access to these two systems.
APPENDIX B – SAMPLE CASEDB CASE

Case 13315712-fromCaseDBRealTimeServer@4/11/200612:34:53AM

CaseCreateDate: 6/28/2005 12:00:41 PM
SiteNumber: Xxxxxxxxx
SiteName: (commercial in confidence)
CaseSNNumber: Xxxxxxxxx
CaseCompletionCode: Closed
CaseText

DialHome
dh1

***NOTES06/28/2005:00:44[06/28/2005:00:44US-GA]ruser0

PRODUCTAWIN_SERIAL_NUMBER=xxxxxxxxxx
PRODUCTAWIN_SITE_NAME=xxxxxxxxxx
PRODUCTA656
RemoteConnect=ProductAIP3.10.03
RemoteControl=ProductARemote4.02.01
Director02a(DA)
SympDateTimeCountDV
119.FF56.0006/28/200510:57:5306/28/200510:57:531
Director06a(FA)
SympDateTimeCountDV
119.FF56.0006/28/200510:57:3806/28/200510:57:381
Director07a(RE)
SympDateTimeCountDV
119.FF56.0006/28/200510:57:3806/28/200510:57:381
Director08a(FA)
SympDateTimeCountDV
119.FF56.0006/28/200510:57:4006/28/200510:57:401
Director09a(FA)
SympDateTimeCountDV
119.FF56.0006/28/200510:57:4006/28/200510:57:401
Director10a(RE)
SympDateTimeCountDV
119.FF56.0006/28/200510:57:3906/28/200510:57:391
Director11a(FA)
SympDateTimeCountDV
119.FF56.0006/28/200510:57:3906/28/200510:57:391
Director15a(DA)

198 Aliases have been used in this sample case to protect the commercial confidentiality of HTG.
<table>
<thead>
<tr>
<th>SymptomDateTimeCountDV</th>
<th>219.680B.5006/28/200510:57:34/28/200510:57:3413FFF</th>
</tr>
</thead>
<tbody>
<tr>
<td>119.FF56.0006/28/200510:57:38/28/200510:57:3813FFF</td>
<td></td>
</tr>
<tr>
<td>Director16a(DA)</td>
<td></td>
</tr>
<tr>
<td>SymptomDateTimeCountDV</td>
<td>219.680B.5006/28/200510:57:35/28/200510:57:3513FFF</td>
</tr>
<tr>
<td>119.FF56.0006/28/200510:57:38/28/200510:57:3813FFF</td>
<td></td>
</tr>
<tr>
<td>Director02b(DA)</td>
<td></td>
</tr>
<tr>
<td>SymptomDateTimeCountDV</td>
<td>219.680B.5006/28/200510:57:33/28/200510:57:3313FFF</td>
</tr>
<tr>
<td>119.FF56.0006/28/200510:57:37/28/200510:57:3713FFF</td>
<td></td>
</tr>
<tr>
<td>Director08b(FA)</td>
<td></td>
</tr>
<tr>
<td>119.FF56.0006/28/200510:57:37/28/200510:57:3713FFF</td>
<td></td>
</tr>
<tr>
<td>Director11b(FA)</td>
<td></td>
</tr>
<tr>
<td>SymptomDateTimeCountDVCH</td>
<td>219.680B.5006/28/200510:57:33/28/200510:57:3313FFF</td>
</tr>
<tr>
<td>119.FF56.0006/28/200510:57:37/28/200510:57:3713FFF</td>
<td></td>
</tr>
<tr>
<td>Director15b(DA)</td>
<td></td>
</tr>
<tr>
<td>SymptomDateTimeCountDV</td>
<td>219.680B.5006/28/200510:57:33/28/200510:57:3313FFF</td>
</tr>
<tr>
<td>119.FF56.0006/28/200510:57:37/28/200510:57:3713FFF</td>
<td></td>
</tr>
<tr>
<td>Director16b(DA)</td>
<td></td>
</tr>
<tr>
<td>SymptomDateTimeCountDV</td>
<td>219.680B.5006/28/200510:57:33/28/200510:57:3313FFF</td>
</tr>
<tr>
<td>119.FF56.0006/28/200510:57:37/28/200510:57:3713FFF</td>
<td></td>
</tr>
<tr>
<td>Director06c(DA)</td>
<td></td>
</tr>
<tr>
<td>119.FF56.0006/28/200510:57:37/28/200510:57:3713FFF</td>
<td></td>
</tr>
<tr>
<td>Director02c(DA)</td>
<td></td>
</tr>
<tr>
<td>SymptomDateTimeCountDV</td>
<td>219.680B.5006/28/200510:57:33/28/200510:57:3313FFF</td>
</tr>
<tr>
<td>119.FF56.0006/28/200510:57:37/28/200510:57:3713FFF</td>
<td></td>
</tr>
<tr>
<td>Director08c(FA)</td>
<td></td>
</tr>
<tr>
<td>119.FF56.0006/28/200510:57:37/28/200510:57:3713FFF</td>
<td></td>
</tr>
<tr>
<td>Director06c(FA)</td>
<td></td>
</tr>
</tbody>
</table>


AdditionalDialHome

dh2
7***DIALProduced***
PRODUCTAWIN_SERIAL_NUMBER=xxxxxxxxxx
PRODUCTAWIN_SITE_NAME=xxxxxxxxxx
**PRODUCTA656**

RemoteConnect=ProductAIP3.10.03  
RemoteControl=ProductARemote4.02.01  
Director02a(DA)

| SympDateTimeCountDV | 104.E00B.006/28/200511:08:20/26/200511:08:2013FFF  
104.E00B.010/28/200511:08:21/28/200511:08:2113FFF  
104.E00B.020/28/200511:08:210/28/200511:08:2113FFF  
104.E00B.030/28/200511:08:22/26/200511:08:2213FFF |

| Director08a(FA) | SympDateTimeCountDVCH | 104.E00B.006/28/200511:08:20/26/200511:08:2013FFF  
104.E00B.010/28/200511:08:21/28/200511:08:2113FFF  
104.E00B.020/28/200511:08:210/28/200511:08:2113FFF  
104.E00B.030/28/200511:08:22/26/200511:08:2213FFF |

| Director09a(FA) | SympDateTimeCountDVCH | 192.11115.2006/28/200511:08:21/28/200511:08:211 |

| Director10a(RE) | SympDateTimeCountDVCH | 104.E00B.006/28/200511:08:20/26/200511:08:2013FFF  
104.E00B.010/28/200511:08:21/28/200511:08:2113FFF  
104.E00B.020/28/200511:08:210/28/200511:08:2113FFF  
104.E00B.030/28/200511:08:210/28/200511:08:2113FFF |

--- Additional Errors Not Shown in CaseDB Case ---

***STATUSCHANGE06/28/2005:18:02Pmvonscb***

Was dialed in and got disconnected. Modem keeps ringing busy. RMAscript was started for replacement of DA1. Please have CE reboot the modem.

**PersonK**

ext 54733

Paging CE... **PersonA**

Paging CE... **PersonB**

Paging CE... **PersonC**

Putting my queue... **PersonD**

Assigning case to CE **PersonD** as per the notes logged... **PersonE**

dialiniscorrected. **PersonF**

CE **PersonF** called in... T/F to HS per his request.... **PersonG**

please have CE call **PersonF** with a status... **PersonH**

please page CE... **PersonJ** needs replacing... **PersonD**

PN: 123-123-123 ASNC12345678

previous HS dialled in tried to run RMAscript but could not get back into box...

**PersonI**

Person I going to dial in and look at log to see what happened yesterday and report to **PersonJ** when complete... **PersonD**

Waiting for **PersonJ** to tell me to close this case... **PersonD**

dareplaced on another case
APPENDIX C – SYSTEM REQUIREMENTS

C.1 Key Performance Indicators

In the support centre context, success can be measured by the effectiveness and efficiency with which customer problems are handled. Hence the following Key Performance Indicators (KPIs) would be relevant to any proposed trouble-shooting workflow solution:

- Increased customer satisfaction;
- Rapid fault and enquiry resolution times;
- Increased quality and consistency of solutions;
- Reduced duplication of solutions;
- Increased accuracy of solution matching as measured by reduced case revisits;
- Reduced problem incidence;
- Increased solution re-use for repetitive problems;
- Reduced organisational re-learning;
- Increased automation of problem diagnosis and solution matching;
- Increased customer self-service for example via an Interactive Voice Response system or via an Internet self-service kiosk;
- Increased in-line self-learning by support centre staff;
- Reduced staff frustration;
- Increased staff satisfaction; and
- Reduced staff turnover.

C.2 Search and Retrieval

The following list of search and retrieval requirements was created on August 7, 2003:

- A way of systematically gathering symptoms that provides a structured approach to both entering data into and retrieving data from existing knowledge bases. Some sort of template interface may be used that asks the user a series of questions, or presents a series
of dropdown or checkbox selections, to help identify symptoms associated with the case at hand.

- An interface that helps the user narrow down the search results obtained from e.g. SolutionDB, so that only relevant solutions are found for the case on-hand.

- A one-stop-shop for retrieving relevant information from all of the existing knowledge bases.

- A way of rating the success of retrieved solutions, and a system that orders retrieved solutions based on the goodness of fit to the described problem.

- A system that uses ontology of terms to indicate synonyms and thereby improve the search and retrieval mechanism of the knowledge base.

### C.3 Knowledge Acquisition (KA)

The following list of KA requirements was created on August 7, 2003:

- A system whose consistency is continually and actively refined by users confirming, or denying existing conclusions in the knowledge base and hence by users providing feedback that is used to continuously improve the system.

- A failure driven system whose completeness is continually and actively improved by systematically prompting users to provide missing knowledge.

- A structured approach to problem solving that identifies the questions users need to ask (e.g. of engineering) when gaps in the knowledge base are identified.

- A system that is able to retain the state of a particular trouble shooting pathway, for example, that the user has traversed a path, come to an invalid conclusion and rejected it, and is now gathering additional case criteria in their hunt for a relevant solution.

- A system that can provide multiple links to a variety of resources e.g. SolutionDB, Docco\textsuperscript{199}, or web pages, a combination of which may help in solving the case at hand.

- A system that bridges the gap between novice and expert by making expertise public.

\textsuperscript{199} An alias for HTG’s document management system.
• A system that overcomes human overload and consistency problems, providing a second opinion for both novices and experts.

• A system that provides partitions of visibility so that perhaps one day customers will be provided access to certain partitions.

C.4 Customer Satisfaction

Adria and Chowdhury (2002) believe that the ultimate purpose or effect of a call-centre implementation “is to streamline the pathway to information for the customer”. They identify three dimensions of call centre employee skill: responsibility, abstractness and interdependence. They use these three dimensions to argue that call centres should allow decisions to be made as close as possible to the customer, including the employee decisions to add to, revise and work with the corporation’s knowledge base. They identified the case of Sun Life, a group insurance company, that took a team approach involving both front-line workers and technical experts to designing the service delivery operations so that the customer could experience a richer real-time interaction. They also highlighted the case of the Mayo Clinic in Rochester, Minnesota, where physicians post links on the clinic’s intranet to web sites that provide up-to-date and authoritative information about current medical treatments for use by clinic practitioners. For call centres to thrive, they argue that support centre personnel need an adequate amount of autonomy and responsibility, and that support centre staff can have a role in updating and correcting a shared knowledge base200.

C.5 Employee Satisfaction

Support Centres can be difficult environments in which to work. Work may be sporadic, often stressful, and there may be a perceived lack of career progression opportunities. Anecdotally, the staff retention period for Support Centres in the ICT industry is around 18-24 months. Generally speaking, low retention rates coupled with poor knowledge re-use has an obvious human and business impact: operational funds are spent training new-hires using limited training and technical resources while those with experience and knowledge walk out the door.

200 Some of this text appears in (Richards and Vazey, 2005).
In an article to the Irish Sunday Business Post, Pearson (2003) ponders the affect of increased computer intelligence in support centres. He suggests that increased computer intelligence could be used to up-skill people by taking over the more mundane tasks giving them time for the more (fuzzy) knowledge intensive tasks than the computer can provide, or time to deal with those enquiries that are more human contact oriented, or to add the human touch to what would otherwise be just a mechanistic response. The end result would be reduced worker stress, improved job satisfaction, and a more enjoyable customer experience. Workers would be humanised by future technology, not made redundant, up skilled instead of being made obsolete.

### C.6 Usability

Usability extends to system responsiveness, ergonomics, ease-of-use and the intuitiveness of the interface. In his evaluation of PEIRS (described in more detail in section 4.3.1.1 on page 46, Edwards noted that information systems will be rejected if they are non-intuitive or if they add to the workload of busy professionals (Edwards, 1996, pp 229). He noted that improving the users’ control over and understanding of the system will facilitate its uptake into routine practice (pp 244 - 246).

If critical mass of both content and users cannot be achieved, then the system will be a failure (Stenmark and Lindgen, 2003). Achieving this critical mass relies not only on a technical solution but also on a solution that addresses the human issues. These issues include the need to provide a strategy that supports easy maintenance that can be performed by multiple staff with domain expertise rather than a single knowledge engineer. The approach needs to be intuitive and easy to learn due to the high turnover of staff within the support centre.

### C.7 Local Concordance

Through the experience of the SCRDR pathology system known as PEIRS (described in detail in section 4.3.1.1 on page 46), Edwards found that pathologists were mistrustful of and resistant to “black box” diagnostic systems, and that local concordance and attention to local work practices is critical if human experts are to accept decision support systems into routine

---

201 http://archives.tcm.ie/businesspost/2003/06/01/story822581121.asp

202 Some of this text appears in (Vazey and Richards, 2006a).
practice (Edwards, 1996, pp v, vii, 18, 41, 46, 48-49, 123, 174). In support of SCRDR and MCRDR, Edwards determined that when knowledge acquisition tasks are minor, experts achieve much greater control over the knowledge and that control, in turn, enables a far greater degree of concordance with local expertise and custom. Hence a further objective is to support easy knowledge acquisition, and hence local concordance with front-line trouble-shooters.

C.8 Workflow Integration

Edwards has previously noted that a degree of process re-engineering is required for system success (Edwards, 1996, p231). The system would need to provide enough incentive to ensure that staff will use it, benefit from it, enjoy it, and maintain it. In other words, the system needs to afford user participation (Stenmark and Lindgen 2003).203 Acceptance of the system would largely depend on successful workflow integration and possibly some workflow redesign where inefficiencies currently exist.

C.9 Feedback and Collaboration

Incoming problems are often solved via reference to a range of materials, including databases and documents from other vendors, and often involve a chain of people who may contribute to one or more (possible) solutions. Therefore feedback and collaboration would be important. Acceptance of a new knowledge-based system would hinge on whether the knowledge it contained was deemed to be accessible, valuable and credible204.

C.10 Organisational Learning and Continuous Improvement

Communities of Practice (CoP) is a term coined by Wenger and Lave (Wenger, 1998) that refers to the way in which people naturally work together. CoP recognises the power of informal peer networks, their creativity and resourcefulness in solving problems, and their ability to create easier ways to reach their goals. Support Centre personnel are continuously exposed to common classes of problems, and they share a desperate need to know what each

---

203 Some of this text appears in (Vazey and Richards, 2006a).

204 Some of this text appears in (Vazey and Richards, 2006b).
other knows in regard to solving new / repetitive / time consuming / or technically challenging problems.

Tsoukas and Vladimirou (2001) reviewed the flow of organisational knowledge within a customer care call centre for Panafon, Greece’s leading mobile phone operator in 2001 and observed that despite the Panafon call centre not being a knowledge-intensive environment, and despite the employee perception that answers to 95% of the questions asked were available “somewhere” in the computer system, several operators were observed constructing their own personal information systems, which contained photocopies of the relevant corporate manuals plus personal notes. In other words, even for a support centre handling problems with low complexity, alongside the formal organisational knowledge there existed an informal knowledge that was generated in action, and which represented the heuristic knowledge residing both in the individual’s minds and in the stories shared in their communities of practice.\(^{205}\)

Basing their analysis on Polyani’s insight concerning the personal character of knowledge, and Wittgenstein’s claim that all knowledge is in a fundamental way collective, Tsoukas and Vladimirou (2001) concluded that knowledge management is the dynamic process of turning an unreflective practice into a reflective one by elucidating the rules guiding the activities of the practice, by helping give a particular shape to the collective understandings, and by facilitating the emergence of heuristic knowledge. They also concluded that it was both feasible and desirable to capture this heuristic knowledge and through casting it into propositional statements, to turn it into organisational knowledge. While the abstract generalisations would be incomplete to capture the totality of organisational knowledge, they concluded that:

“The more propositional statements and collective understandings become instrumentalised (in Polanyi’s sense of the term); and the more new experiences are reflectively processed (both individually and collectively) and then gradually driven into subsidiary awareness, the more organisational members dwell in all of them, and the more able they become to concentrate on new experiences, on the operational plane” (Tsoukas and Vladimirou 2001, p. 983).

\(^{205}\) Some of this text appears in (Richards & Vazey 2005).
In accordance with Polyani’s observations that knowledge always contains a personal element, they noted that an improvisational element would still be required for each interpretation of the collective organisational knowledge, and that it was the dialectic between the general and the particular, that gave organisational knowledge its dynamism. Finally, they concluded that:

“\textit{The effective management of organisational knowledge requires that the relationship between propositional and heuristic knowledge be a two-way street: while propositional knowledge is fed into organisational members and is instrumentalised through application (thus becoming tacit), heuristic knowledge needs to be formalised (to the extent that this is possible) and made organisationally available}” (Tsoukas and Vladimirou 2001, p. 991).

The benefit of this type of organisational learning to business stakeholders is a better bottom line; the benefit to customers is a more responsive and reliable service; and the benefit to staff is the empowerment that they may feel in being able to contribute to a common knowledge base, the learning that they will receive from the common knowledge base, and the elimination of menial work researching and solving the more repetitive types of problems.

In summary, a system that promotes organisational learning would also support continuous organisational improvement.

\textbf{C.11 Other Requirements}

The following list of other requirements emerged over the course of the project:

- The system should be accessible by each and every user who can make a positive contribution. The solution must support easy maintenance that can be performed by multiple call-centre staff with domain expertise rather than a single knowledge engineer. The content must be able to be entered and maintained with minimal editorial hardship. As well, the system should be able to rely on people’s inherent self-interest for its maintenance, that is, that people will maintain the system because it benefits them to do so.

- The approach needs to be intuitive and easy to learn due to the high turnover of staff in the Level One group.

- Staff incentive structures may need to be modified to support the workflow.
Appendix C – System Requirements

- Solutions need to contain a solid description of the problem, including the problem symptoms and a description of exactly what does and doesn’t work. Bad solutions can be worse than no solution as they may direct field personnel to implement flawed solutions that actually worsen the customer’s position.

- Many businesses are encouraging their customers to seek their own help and solve their own software and/or hardware problems via Internet customer service portals. HTG is also keen to move in this direction and has already established an e-service solution. Any proposed future solution would need to also fit with the self-service web-kiosk paradigm offered by HTG’s existing Internet help-desk portal.
APPENDIX D – CONCEPTS FROM THE LITERATURE

D.1 Information Retrieval (IR)

Information Retrieval traditionally involves retrieving desired information from textual data (Dunham, 2003, p26). I prefer to expand the definition of Information Retrieval (IR) to include the retrieval of non-textual data, for example hyperlinks to intranet or extranet websites, where hyperlinks can point to any network accessible object for example a sound file, image file, video file, binary or text file, or other document of otherwise arbitrary format.

IR has its historical roots in the effective use of libraries. The IR task basically reduces to a classification task because the set of documents in the library is divided into classes based on the keywords involved. The effectiveness of the IR task is often measured by precision and recall (Dunham, 2003, p26) where:

\[
\text{precision} = \frac{\text{number of relevant and retrieved documents}}{\text{number of documents actually retrieved}}
\]

\[
\text{recall} = \frac{\text{number of relevant and retrieved documents}}{\text{number of documents that were actually relevant (but not necessarily retrieved)}}
\]

The objective of IR is to maximise both the precision and recall for the user.

The inverse document frequency (IDF) is often used in IR as a similarity measure. It assumes that the importance of a keyword in calculating similarity measures is inversely proportional to the number of documents that contain it. In other words, the precision and recall of IR will be enhanced by the uniqueness of the search keywords used.

D.2 Decision Support Systems (DSS)

Decision Support Systems (DSS) are one example of IR systems. DSS are comprehensive computer systems and related tools that assist people (and often managers) in making decisions and solving problems (Dunham, 2003, p28). They can incorporate Executive Information Systems (EIS) and Executive Support Systems (ESS), and are part of a suite of systems known as Management Information Systems (MIS). In DSS, information is retrieved for the purpose of making decisions.

Many, but not all DSS incorporate knowledge discovery in databases.
Knowledge Discovery in Databases (KDD)

Knowledge Discovery in Databases (KDD) is the pursuit of useful information and patterns in data, and Data Mining is the use of algorithms in that pursuit (Dunham, 2003, p9). The two terms are sometimes used interchangeably, however KDD involves some combination of Data Mining or Knowledge Acquisition where:

- Data Mining (DM) involves the use of clever data discovery algorithms, and
- Knowledge Acquisition (KA) includes expert human review of the extracted data together with manual fine-tuning of the relevant models and algorithms.

(Adapted from Dunham, 2003, p9.) DM and KA are described in further detail later on in this Appendix.

Artificial Intelligence (AI)

The term Artificial Intelligence (AI) was coined in a seminar in 1956 at Dartmouth College (Oz, 2004, p514). AI researchers try to emulate the human mind in machines, hence in this context, intelligence is the ability to acquire and apply knowledge, to think, to reason, and to learn. It includes making associations between a previous experience and a new situation, drawing systematic conclusions, quickly adopting new ways to solve problems, being able to separate what is important from what is not in solving a problem, and determining what tools can or cannot help in handling a complex situation.

Machine Learning (ML)

Machine Learning (ML) is an area of AI that examines how to write programs that can learn (Dunham, 2003, p43). ML incorporates supervised and unsupervised learning.

- A supervised approach learns by domain knowledge, by example, and via feedback. It takes a training set of data, plus its correct answers, and develops and refines a model of the data (such as a graphical structure like a neural network or a decision tree) based on these examples. The model can then be used to make future predictions or recommendations, including classifications.

- With unsupervised learning, data exist but the expected outcome of applying the model to the data is unknown. Unsupervised learning may involve techniques such as genetic algorithms or clustering.
ML has a focus on learning. It concentrates on tasks that may be difficult for humans to do, or on how to develop learning techniques that can mimic human behaviour.

In contrast to ML, rather than trying to take their place, DM has a focus on retrieving information that will be useful for humans (Dunham, 2003, p44).

Many of the algorithms introduced by AI and ML are now used in the DM task.

**D.6 Data Mining (DM)**

Data mining\(^{206}\) models can be classified into techniques that are descriptive versus those that are predictive (Durham, 2003):

- Descriptive models are aimed at identifying patterns or relationships in data so as to understand or express the nature of the data. Methods used to create descriptive models include clustering, summarisation, association rules, and sequence discovery.

- Predictive models allow future values to be predicted, for example using historical or correlated data. Methods used to create predictive models include classification; pattern recognition; time series analysis; and / or regression analysis.

In addition, statistics are often an important part of validating emerging models, and ensuring that future predictions are statistically significant.

Descriptive techniques and models are suited to unsupervised learning, whereas predictive techniques and models are suited to supervised learning tasks.

**Descriptive DM Models**

Clustering is an unsupervised learning technique that is used to identify like data (Dunham, 2003). Clusters, categories or classifications are discovered based on similarities and differences between the data. Bayesian classification or other statistical clustering techniques may be useful. A domain expert is often required to interpret the meaning of the derived clusters. Summarisation is one application of clustering algorithms – it is a process of generalising the data in order to provide a “helicopter view” or top-level summary of it.

\(^{206}\) see the ACM’s KDD special interest group at [www.acm.org/sigkdd](http://www.acm.org/sigkdd), and also the vendor-led data-mining group at [www.dmg.org](http://www.dmg.org)
Association rules look for co-occurrences or correlations by examining the behaviour of attributes across the data set. The resulting rules may be observational rather than causal. Sequence discovery is a subset of the association rule discovery task and seeks to discover association rules on the basis of time.

**Predictive DM Models**

Classification is a supervised learning technique where data is classified into predefined categories (Dunham, 2003). Pattern recognition can be considered as a subtask of classification in that it seeks to discover and define the rules that determine when a particular classification is appropriate.

Time series analysis is used to determine correlations, to confirm models, and to predict future values. Regression analysis is one tool that can be used for time series or other analysis – for example in linear regression analysis we seek to fit a linear model to the data and minimise the error of the fit by changing the coefficients of the linear model.

**D.7 Popular AI, ML and DM Techniques**

**Decision Trees**

Decision trees are based on the “20 questions” game that kids play. Decision trees only have one input node, which is the root of the tree. Each node contains a question, and the arcs emanating from a node represent the possible answers to that node’s question. In a binary decision tree, the optimal strategy is to divide the search space into two equal parts at every node and hence converge to a solution with the fewest number of question-answer pairs (adapted from Dunham, 2003, p58).

Decision trees assume that the outcome of the decision process is one, as opposed to multiple conclusions. They offer a divide and conquer technique that can be used in classification, clustering and prediction tasks.

**Neural Networks (NNs)**

The NN method takes a set of training examples, and if available - the desired outcomes (in the case of supervised learning), together with an initial network graph, and applies a learning algorithm to tune the network graph so as to create a predictive model of the data (adapted from Dunham, 2003). NNs can have a long training time and therefore may not be
appropriate for real-time applications. A typical problem with NNs for open learning problems is that there are too many unknowns, or too many degrees of freedom, so convergence to a useable model never occurs. In practice much human input is required to eliminate the extra degrees of freedom. NNs have been used with some success in speech recognition applications.

Neural Networks (NNs) are structured as a directed graph with many processing element (nodes), and interconnections (arcs). The main idea with NNs is to model the available data. NNs are suited to numeric input data.

NNs have one input node for each attribute, and one output node for each possible classification. There can be any number of hidden layers between the input and output nodes (but typically just one or two), which are comprised of nodes that take as their input the weighted sum of the outputs from preceding nodes and provide some signal as their output which multiplies the weighted and summed inputs with some activation function that may be one of several different forms, for example: threshold (step), sigmoid, gaussian (impulse), linear, or hyperbolic tangent; where the range of the activation function may be unipolar with values [0, 1] or bipolar with values [-1, 1] (adapted from Dunham, 2003 p62).

One of the key differences between the NN and decision tree structure, is that a NN can generate multiple possible outcomes each with its own statistical likelihood of relevance. As well, in the NN structure, successor child nodes may have more than one preceding parent node and hence there can be multiple paths through the network to the same outcome. This structure is particularly relevant for the types of classification problems observed at the support centre.

In AI, NNs can be feed-forward only, or they can include feedback where links are provided back to earlier layers in the network. Hence although the NN graph is directed, it does not need to be acyclic. This has parallels with the idea of recursive question-answering in decision trees.

**Genetic Algorithms (GAs)**

Genetic Algorithms (GAs) are examples of evolutionary computing methods (adapted from Dunham, 2003). GAs take a model and some initial solutions, and through the processes of gene-allele (attribute-value) mutation and crossover between individual candidates within the test population, the GA hunts for the best individuals or exemplars that fit the model. One of
the requirements for the GA is that the user defines a fitness function that can specify the 
goodness of fit of a particular individual to the model. As well, suitable mutation and 
crossover algorithms, and suitable seed candidates are required.

D.8 Problems with AI, ML and DM Techniques

In general, some of the problems that arise when traditional AI, ML and DM techniques are 
applied to the KDD task are that they have limited capacity to deal with the complexities of 
real-life databases (adapted and augmented from Dunham (2003, p44 and p14), including:

- their enormous size – possibly millions of records;
- complex data types – not just numeric and / or simple keywords (for example 
multimedia data);
- high dimensionality – many different attributes;
- complex data relationships – attribute correlations and dependencies;
- noisy, erroneous, irrelevant or missing data together with data outliers; and
- constantly changing and hence dynamic data.

As well, the following broader issues need consideration:

- suitable training data is required which may need scrubbing;
- expert input is required to avoid over-or under-fitting of the learnt model(s);
- results may still require interpretation;
- to assist with comprehension, techniques may be required to help experts visualise and 
better understand the results;
- experts have different opinions about what the most useful and / or important data is;
- results may be needed in real-time;
- the resulting system will need to be integrated with other (legacy) systems to support 
the normal organisation workflow; and
- significant organisational change may be required to effectively use the system and 
make use of its results.
APPENDIX E – TACIT KNOWLEDGE

Gourlay (2002) provides an overview of tacit knowledge and refers to Polanyi’s (1958) original work on tacit knowing and Nonaka and Taceuchi’s (1995) introduction of the term to the knowledge management field. Gourlay (2002, p2) also refers to the work of Baumard (1999) who argued that tacit knowledge is inexpressible.

Tacit knowledge can be thought of as analogous with a continuous analogue signal representing all the things that we know but which we either can’t or don’t articulate. One of the reasons why we can’t articulate some of this knowledge is that we simply don’t have time, capacity, or a format in which to express the knowledge. There are infinite207 shades of tacit knowledge between each of the discrete concepts that we choose to express. In contrast, explicit knowledge can be thought of as all the things that we know that we choose to articulate. In other words, our explicit knowledge represents a digital sampling of our analogue tacit knowledge. Just as the colours we choose to articulate e.g. red, green, blue, yellow are digital or quantum representations of our continuous, analogue and infinite experience of colour.

The interesting thing about humans and hence the trouble-shooting process is that the nature and structure of questions asked will determine which knowledge is made explicit in the context of the problem on hand. In many cases it seems that when the question is ready, the answers appear. This reflects the constructive nature of human memory (John Gero208, personal communication, 2006) and is supported by the findings of Coward’s Recommendation Architecture in the field of cognitive science (Coward, 2006).

One of the focuses of this research was therefore to formally prepare questions that challenge a team of experts to share with their team-mates the properties of each incoming unique class of problem that can be used to map out a path that will consistently lead subsequent trouble-shooters to an appropriate solution.

207 This view is supported by Dazeley and Kang (2004, p2) who note that while expert systems can capture the vast majority of knowledge required in some domains: “the potential knowledge needed is actually infinite”.

208 Key centre of Design Computing and Cognition, Sydney University.
Appendix F – Ontologies

APPENDIX F – ONTOLOGIES

Table 30 provides a list of ontology editors from Bao and Honavar (2004) that claim to support collaborative ontology editing.

Table 30: Collaborative Ontology Editors

<table>
<thead>
<tr>
<th>Tool</th>
<th>Base Language</th>
<th>Import/Export</th>
<th>More Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>IODE</td>
<td></td>
<td>XML, RDF(S), DAML+OIL, OWL</td>
<td><a href="http://www.ontologyworks.com/">http://www.ontologyworks.com/</a></td>
</tr>
<tr>
<td>LinkFactory Workbench</td>
<td></td>
<td>DAML+OIL, XML, KIF</td>
<td><a href="http://kaon.semanticweb.org/">http://kaon.semanticweb.org/</a></td>
</tr>
<tr>
<td>Onto-Builder</td>
<td></td>
<td>RDF(S), F-Logic, DAML+OIL, RDF</td>
<td><a href="http://www.land.de/">http://www.land.de/</a></td>
</tr>
<tr>
<td>Ontosaurus</td>
<td></td>
<td>DAML/RDF, CGI, DAML+OIL, RDF, X-CARIN, FLogic, Prolog, XML</td>
<td><a href="http://www.toplingue.com/">http://www.toplingue.com/</a></td>
</tr>
<tr>
<td>OpenKaoMe</td>
<td></td>
<td>RDF, RDFS, GXL, Ontolingua, OIL</td>
<td><a href="http://megunesia.int.gu.edu.au/">http://megunesia.int.gu.edu.au/</a> phmartin/WebKB/</td>
</tr>
<tr>
<td>WebKB</td>
<td></td>
<td></td>
<td><a href="http://delicious.dia.f.upm.es/webODE/">http://delicious.dia.f.upm.es/webODE/</a></td>
</tr>
<tr>
<td>WebODE</td>
<td></td>
<td></td>
<td><a href="http://kmi.open.ac.uk/projects/webonto/">http://kmi.open.ac.uk/projects/webonto/</a></td>
</tr>
<tr>
<td>WebOnto</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Reproduced from Bao and Honavar (2004, p9) with the kind permission of Vasant Honavar.

As noted by Bao and Honavar (2004, pp 1, 8), many of the ontology editors shown in Table 30 provide concurrent access control with transaction oriented locking and in some cases even rollback, however:

“relatively little attention has been paid to the development of principled approaches and tools for collaborative ontology building”. “None of the existing ontology editors, to the best of our knowledge, provides principled approaches for manipulating independently developed, semantically heterogenous ontology modules or for reconciling logical inconsistencies between such modules”.

“There is an urgent need for principled approaches and flexible tools for allowing individuals to collaboratively build, refine, and integrate existing ontologies as needed in specific contexts or for specific applications”.

Duineveld et al. (1999)209 have also provided a comparative study of ontological engineering tools, including Ontolingua, WebOnto, ProtegeWin, OntoSaurus, and ODE. Their findings are reproduced in Table 31 below.

---

Reproduced from Duineveld et al. (1999)

Duineveld et al. found that in 1999, Ontolingua was the only tool to support the full synchronous editing (3.1 in the table) of the ontology by multiple users. It allowed different users from all over the world to work together in constructing a single ontology, and to have their software client instantly updated whenever someone else made a change. However, as with all the other ontology editors surveyed, users were not directly notified of the changes and when a change was discovered (3.4 in the table), it was not possible to find out which user had made the change.

Interestingly they found that WebOnto, Protégé and ODE were best suited for the early conceptualisation and formalisation phase of the ontology development, whereas Ontolingua and Ontosaurus were better suited to the later phases where only relatively small revisions were required. It was felt that these latter two might also give the users better support in creating complex ontologies.

Table 31: Wondertools: A summary of the results. + means positive, 0 means reasonable, - means negative, and NA is “not applicable”.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Ontolingua</th>
<th>WebOnto</th>
<th>Protégé</th>
<th>Ontosaurus</th>
<th>ODE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Interface clarity</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1. Interface consistency</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>1.3 Speed of updating</td>
<td>-</td>
<td>0</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1.4 Overview</td>
<td>0</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>1.5 Meaning of commands</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>1.6 Identifiability of changes</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.7 Stability</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>1.8 Local installation</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes/No</td>
<td>Yes</td>
</tr>
<tr>
<td>1.9 Help-system</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

2. Ontology

2.1 Multiple inheritance       | Yes        | Yes     | Yes     | Yes        | Yes |
2.2 Decomposition types        | +          | +       | +       | +          | +   |
2.3.1 Consistency checking     | +          | +       | +       | +          | +   |
2.3.2 Level of checking        | ?          | 0       | 0       | ?          | +   |
2.4 Example ontologies          | +          | +       | 0       | +          | 0   |
2.5 Reusable ontologies         | +          | +       | -       | +          | +   |
2.6 High-level primitives      | +          | +       | -       | +          | +   |
2.7 Ontological help            | -          | -       | -       | -          | -   |

3. Computation

3.1 Synchronous editing        | +          | +       | -       | +          | +   |
3.2 Ontology locking            | +          | +       | -       | +          | +   |
3.3 Browsing when locked        | +          | +       | NA      | +          | NA  |
3.4 Change recognition          | -          | -       | -       | -          | -   |
3.5 Export facilities           | +          | -       | -       | +          | 0   |
3.6 Import facilities           | +          | -       | -       | +          | +   |
Other interesting ontology editors and databases include OpenCYC\textsuperscript{210}, Wordnet\textsuperscript{211}, Jena\textsuperscript{212}, OilEd, and Ontolinguia.

A further survey of ontology editing tools is available at:

\url{http://www.xml.com/2002/11/06/Ontology_Editor_Survey.html} (Michael Denny, 2002).

\textsuperscript{210} \url{http://www.opencyc.org/faq/opencyc_faq}

\textsuperscript{211} see \url{http://wordnet.princeton.edu} and \url{http://en.wikipedia.org/wiki/Wordnet}

\textsuperscript{212} \url{http://jena.sourceforge.net/index.html}
APPENDIX G – TRANSCRIPT OF PKS INTERVIEW

I arranged to meet with Lindsay Peters, the Chief Technology Officer of Pacific Knowledge Systems (PKS), at their Australian Technology Park (ATP) Eveleigh / Redfern offices on June 16th, 2005. Dr. Debbie Richards and I spoke with Lindsay Peters and these are the notes from that interview. These notes were reviewed by Lindsay Peters after the interview, and updated according to his feedback to ensure the accuracy of the findings reported herein.


**G.1 Scope**

At this time, PKS have 7 paying customers using their MCRDR-based LabWizard knowledge management product. Some of the customers use LabWizard at multiple lab sites, for example in different states.

Many of these customers are in the pathology domain, for example building knowledge bases to assist with interpretations for diabetes and bone density, however they also have a customer, Bayer crop-science, working in the area of DNA testing of soil.

In sum, LabWizard generates upward of 600,000 interpretations per month across the PKS client base.

90% of all of these reports are auto-validated i.e. they require no human review since the knowledge base has become so mature that the users are confident of its accuracy in 90% of cases.

**G.2 New Clients**

Further, PKS have 3 trial sites generating 300,000 interpretations per month, including a 6-month trial site in Holland.

Another trial site is expected to come online shortly in Vancouver, Canada.

PKS is also undergoing major discussions with a bank to provide a LabWizard style Help Desk application, in particular to support card services and loan approvals. The idea is to provide a questionnaire that uses the rule base to provide more targeted questions based on the user’s previous answers.
PKS has also been working on a system to provide automated real estate valuations on property to assist with loan approvals.

Most new clients are offered a 45-90 day trial of LabWizard.

**G.3 History**

LabWizard versions 3, 4, and 5 represent the progression from prototype, to trial version and eventually full version. There have been 50 minor releases of the software over the last 5 years. Since 2002 there has been a mature Java version of LabWizard (version 5).

3 years ago, PKS downsized to just 4 staff – two software engineers, a pathologist, and a CEO. Today they feel that they have turned the corner and are entering a growth phase. They expect to be recruiting more software engineers in 2H2005.

**G.4 Size of the knowledge bases**

Most of the PKS knowledge bases are in excess of 1000 rules and one of them is in excess of 10,000 rules.

Much of the domain can be covered in the first two-week period by building the first few hundred rules. A rule base needs 200-300 rules before it starts to become useful and hence usable.

Clients are typically adding 40-50 rules per month in mature knowledge bases. The knowledge is by no means static in pathology – the rule tree does not reflect the blind application of textbook guidelines but rather the ongoing practical learning of pathologists.

As with any new-starter, training the knowledgebase becomes a bottleneck until a sufficient level of accuracy and hence auto-interpretation is reached.

**G.5 Level of reuse**

A surprising number of cases will require new rules. Despite this, combinations of rules and hence their conclusions can be unique; the system can significantly assist with guided trouble-
shooting through recursive question answering; and higher level information can be just as useful for interpretations that don’t have more specific conclusions.

**G.6 Comment on IP**

Ownership of the IP in the knowledge base is the only reason to separate the knowledge base – there is no technical or performance related reason to separate them.

**G.7 Performance of the knowledge bases**

Labwizard operating on a standard windows server can process 50 cases per second; and 4 transactions per second. Most of the reports are done in batches for example 500 reports at a time.

Independent of the size of the knowledge base, it still takes less than 1 minute to add a rule.

**G.8 Pricing Model**

For most customers in the pathology domain, PKS charges per patient. There are on average two interpretations per patient. This pricing model parallels the manner in which the government charges pathology labs. As well, pathology consumables are typically costed per patient.

For non-pathology customers such as Bayer crop-science the charge is per interpretation i.e. report.

All customers pay a yearly maintenance and license fee.

Some customers don’t pay per report but rather pay an increased license fee only.

The idea with the per-report charge is that it provides for a lower entry cost for customers.

**G.9 Training**

PKS offers training on-site for new customers including a 1.5 day training session for pathologists.

---

213 As noted in section 2.5.2 (page 24), the importance of being able to fall-back on a layer of higher-level general knowledge when more detailed and specific knowledge is missing is also expressed by Dazeley and Kang (2004, p2).
This training is followed by 1-2 hours each day while building the knowledge base over the next two weeks.

**G.10 Collaborative Knowledge Building**

*At this stage only one person builds the rules.* There is more emphasis on end-user validation. Users must validate if the interpretation is not set to auto-validate. Only a handful of users are ever rule-builders. Most users are limited to providing validation feedback in plain English.

At this stage *they don’t allow two separate people to add rules to the same knowledge base simultaneously.* As well, each user can only view one case at a time – cases are queued, but multiple users can validate the same queue of cases simultaneously

*Users can’t reopen a closed case,* that is, once the interpretation has been approved and sent back to the interface information system.

The knowledge base becomes a facilitative forum for resolving conflicts and for sharing information. *Experts can go through the cases one by one together and share their interpretations.*

Individuals can build their own knowledge bases if they want to.

**G.11 Quality Standards**

The system assists in delivering high quality standard interpretations – consistency in policy provides end-clients with the perception of quality.

**G.12 Technical features**

In addition to standard MCRDR features, LabWizard provides for:

- Ordering of conclusions.

- *Hiding of multiple identical conclusions instead of duplicate display.*

- Automatic generation of interpretations in the form of letters, for example a bone density report might include when the patient was last examined, the date of the current visit, and a comment on how they have progressed.
• Documentation of the knowledge base – allowing it to be navigable. Users don’t want it to be black box and there can be a requirement to audit the system e.g. “who made this rule?”. Documentation is usually with a specific objective in mind e.g. what symptoms indicate that a patient is diabetic? **The knowledge base needs to have a balance between being unstructured and structured.**

• The contexts and history of each conclusion are saved.

• Each site sends a heartbeat every 15 minutes to confirm the health of the LabWizard server.

• Conclusions can be actions or alerts in the form of code output that the information system can interpret. There is a socket-based connection to the information server.

• Cases are queued for review.

• An “Auditor” feature allows data entry errors to be checked in the lab and flags entry cases in risk of being in error.

• Labwizard includes several text processing features supporting a variety of output formats, as well as parsing input in natural language format. It is a challenge to make the language of the conditions ie the condition syntax as natural as possible e.g. “that at least 3 high glucose readings have been recorded in the last 5 months”. Templates are provided for condition syntax based on the case. Users can also type in free-text conditions.

• Up until now, LabWizard has allowed for only one person creating rules per domain. They are designing a further enhancement that allows the site ID to be appended to the rules so that users may get a different interpretation depending on where the case comes from.

• A process of reconciling site-specific rules with general rules by relaxing the rule conditions has been created.

• **Incorrect rules are never changed since “rules are cheap”. However PKS is looking at relaxing those conditions.**

• Statistics are kept on the accuracy of interpretations and hence the maturity of the knowledge base.
• There is a set of knowledge bases, a set of users, and a server with a queue of cases and their reports.

• *Difference lists are rarely used since most users know what conditions they want to apply without needing to see a difference list.*
CSIRO’s Panoptic project found that beyond the information provided by anchor text, webpage metadata provided by webmasters did not add any useful additional information for the search algorithm (Personal Communication, David Hawking - CSIRO, HCSNet Summer School - Sydney, Dec 13-14 2005). Reasons were thought to include:

1. Conflicts and ambiguities arising from distributing appropriate metadata labels across a top-level home page compared to its dependent home pages.

2. That keywords used in the metadata were almost always already represented in the anchor text, as well as elsewhere e.g. in the page title or the page content.

3. That user queries displayed a much greater variety than was captured in the metadata itself.

4. That many web pages did not contain any metadata; and/or just repeated the same standard string for the website; and/or the use of metadata fields was inconsistent or erroneous as well as hidden and unlikely to be corrected.

As well, the Panoptic project found that user-clicks on the search result pages provided useful feedback information for the search engine, but when combined with the much greater information gleaned from anchor-text, very little incremental benefit was obtained through the user-click data.

The key finding was that as far as search engines are concerned, anchor-text provides by far the most useful correlation with the terms used by real users to find web pages, probably because anchor text:

1. Reflects the relative value of the target;

2. Reflects a diversity of views from multiple webmasters, not just the view of the webmaster of the target site; and

3. Is visible to people navigating to and viewing the web pages (this may have the effect of training other webmasters).

Appendix H – About Anchor Text

APPENDIX H – ABOUT ANCHOR TEXT

For a definition of Anchor Text, please see the glossary (page 277).
APPENDIX I – FOLKSOMONIES

1.1 del.icio.us

http://del.icio.us is a collection of favourite websites. Users can keep links to their favourite articles, blogs, music, restaurant reviews; share favourites with friends, family, and colleagues; and discover new things. The key idea behind the site is to use tags\textsuperscript{215}, which are descriptors that users can assign to any favourite. Users can assign as many tags to a favourite as they like. Users can also gather together sets of tags into bundles\textsuperscript{216} which can act as categories or classifications that provide the union or intersection of sets of tagged items.

Tagging can be a whole lot easier and more flexible than fitting information into preconceived categories. User’s don’t have to rely on the designer of the system to provide them with a category. They just make up tags as they need them. Apart from being great for organizing personal data, when someone else posts related content with the same tags users begin building a collaborative repository of related information, driven by personal interests and creative organization. The most popular tags are displayed in a tag cloud\textsuperscript{217} that provides an ordered list of tags in which the size of the tag represents its popularity.

For instance, to view del.icio.us user's music favourites, visit http://del.icio.us/tag/music.

1.2 flickr.com

http://flickr.com/ is similar in concept to http://del.icio.us but provides a digital photo library for individuals and their invited communities. Users set privacy levels on each of their photos to determine who can see them and comment on them. Users and their invited visitors can assign as many tags\textsuperscript{218} as they wish to each photo. Tags, which are like keyword or category labels, help users find photos that have something in common. Flickr provides for personal,
popular\textsuperscript{219}, group, and standardised tag usage. It also allows tag clusters of photos to be displayed. Flickr clusters are similar in concept to the categories or bundles provided by del.icio.us, however in flickr they are system determined rather than user defined.

\textsuperscript{219} The 150 most popular flickr tags are shown at http://flickr.com/photos/tags/
APPENDIX J – WIKIPEDIA VERSUS BRITANNICA

The English version of Wikipedia now has a huge size advantage over Encyclopaedia Britannica (930,000 articles compared to 120,000, and 340 million words compared to 55 million)\(^{220}\) and it is FREE! Wikipedia’s English version has more than 45,000 registered users, and users added about 1,500 new articles every day of October 2005. Wikipedia has become the 30th most visited website globally (Alexa\(^{221}\), Jan 2006). It exceeds Encyclopaedia Britannica in elasticity and responsiveness, and surpasses it as a definitive online knowledge resource for the modern world.

One of the biggest challenges to Britannica\(^ {222}\) is the ease with which people can find information at Wikipedia, and online in general. Many people prefer to find information with the help of a search engine\(^{223}\).

Britannica’s main edge over Wikipedia, as a professionally edited resource, is its slightly more consistent reliability: An expert-led investigation carried out by the journal Nature\(^{224}\) (Giles, 2005) used peer review from the scientific community to compare Wikipedia and Britannica's coverage of science. The exercise revealed numerous errors in both encyclopaedias, but among 42 entries tested, the difference in accuracy was not particularly great: the average science entry in Wikipedia contained around four inaccuracies; Britannica, about three. Only eight serious errors, such as misinterpretations of important concepts, were detected in the pairs of articles reviewed, four from each encyclopaedia. Given that Wikipedia entries are less than 5 years old and were freely contributed by the public Internet community, whereas the majority of Britannica entries are more than 10 years old and written by paid professionals; and that Wikipedia has more than 7 times as many entries, it’s a phenomenally good result for Wikipedia.

\(^{220}\) http://en.wikipedia.org/wiki/Britanica


\(^{222}\) http://en.wikipedia.org/wiki/Encyclop%C3%A6dia_Britannica

\(^{223}\) This also helps to explain the enormous successes of businesses like http://www.amazon.com and http://www.ebay.com.

Appendix J – Wikipedia versus Britannica

According to Giles (2005), Wales plans to introduce a 'stable' version of each entry at Wikipedia such that once an article reaches a specific quality threshold it will be marked as stable. Further edits will be made to a separate 'live' version that would replace the stable version when deemed to be a significant improvement. One method for determining that threshold, where users rate article quality, will be trialled in 2006.
APPENDIX K – SEMANTIC WEB LANGUAGES

The aim of the Semantic Web\(^{225}\) is to provide a common framework for data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by the World Wide Web Consortium\(^{226}\) (W3C) with participation from a large number of researchers and industrial partners.

Research is based on the Resource Description Framework (RDF), which integrates a variety of applications using XML for syntax, and Uniform Resource Identifiers (URIs) for naming\(^{227}\). In Feb 2004, the W3C released the Resource Description Framework (RDF) and the OWL Web Ontology Language (OWL) as W3C Recommendations\(^{228}\). These languages are described briefly below.

**K.1 Resource Description Framework (RDF)**

RDF\(^{229}\) is an XML\(^{230}\) based language for representing information about resources in the World Wide Web. It is based on the idea of identifying things using Internet URIs, and describing resources with sets of triples including the resource, its properties and its property values.

**K.2 Web Ontology Language (OWL)**

OWL is the most recent development in standard ontology languages, endorsed by the W3C to promote the Semantic Web vision\(^{231}\) (source: [http://protege.stanford.edu/overview/protege-owl.html](http://protege.stanford.edu/overview/protege-owl.html), April 2006).

---


\(^{226}\) [http://www.w3.org/](http://www.w3.org/)

\(^{227}\) source: [http://www.w3.org/2001/sw/#spec](http://www.w3.org/2001/sw/#spec), April 2006

\(^{228}\) [http://www.w3.org/2001/sw/#spec](http://www.w3.org/2001/sw/#spec)

\(^{229}\) [http://www.w3.org/TR/rdf-primer/](http://www.w3.org/TR/rdf-primer/)


\(^{231}\) [http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&ref=sciam](http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&ref=sciam)
Appendix K – Semantic Web Languages

OWL builds on RDF and RDF Schema\(^{232}\) and adds more vocabulary for describing properties and classes including relations between classes (e.g. disjointness), cardinality (e.g. "exactly one"), equality, richer typing of properties, characteristics of properties (e.g. symmetry), and enumerated classes.

The OWL Web Ontology Language is designed for use by applications that need to process the content of information instead of just presenting information to humans. OWL facilitates greater machine interpretability of Web content than that supported by XML, RDF, and RDF Schema (RDF-S) by providing additional vocabulary along with a formal semantics. OWL has three increasingly-expressive sublanguages: OWL Lite, OWL DL, and OWL Full. (source: [http://www.w3.org/2004/OWL/](http://www.w3.org/2004/OWL/), April 2006).

OWL extends the Resource Description Framework (RDF) and is derived from the DAML\(^{233}\)+OIL (Defense Advanced Research Projects Agency (DARPA)\(^{234}\) agent markup language together with the Ontology Interchange Language). A review of past knowledge sharing efforts by DARPA is provided in Appendix L (page 421).

\(^{232}\) [http://www.w3.org/TR/rdf-schema/](http://www.w3.org/TR/rdf-schema/)

\(^{233}\) see [http://www.daml.org/about.html](http://www.daml.org/about.html) and [http://pride.daml.org/index.html](http://pride.daml.org/index.html)

APPENDIX L – DARPA KNOWLEDGE SHARING

Patil et al. reported on a major knowledge-sharing initiative sponsored by the US Defense Advanced Research Projects Agency (DARPA), the Air Force Office of Scientific Research (AFOSR), the Corporation for National Research Initiative (NRI), and the National Science Foundation (NSF), to develop the technical infrastructure to support the sharing of knowledge among systems (Patil, 1992). This work was published by Nebel et al. (1998).

The goal of the effort was to develop technologies that would enable researchers to efficiently and effectively develop new systems by selecting components from libraries of reusable modules and assembling them together. The idea was to allow the sharing of declarative knowledge, problem solving techniques, and reasoning services so that through economies of scale larger systems could be built more cheaply and reliably. The reusable modules would benefit from refinements made possible by their extensive use. Hence greater ubiquity would be achieved and refinement of the reusable modules would be perpetuated.

Four key areas were identified by the knowing sharing initiative: (1) mechanisms for translation between knowledge bases represented in different languages; (2) common versions of languages and reasoning modules within families of representational paradigm; (3) protocols for communication between separate knowledge based modules, as well as between knowledge-based systems and databases; and (4) libraries of ontologies, i.e. pre-fabricated foundations for application specific knowledge bases in a particular topic area.

In this Appendix, the outcomes of each of the first three of these areas are described. Ontologies have been previously referred to in section 3.4 on page 34. The purpose of including this Appendix is just to highlight the broader milieu in which KBSs are being developed and in particular, the importance of considering issues of knowledge interchange, accessibility and reuse.

L.1 An Interlingua for Knowledge Interchange

A Knowledge Interchange Format (KIF)235 was developed to provide an intermediary knowledge interchange language with sufficient expressive power to allow translation into

235 The draft American National Standard for KIF is located at http://logic.stanford.edu/kif/dpans.html
and out of different KBS representation languages. KIF version 3.0 was an extended version of first order predicate logic and was comprised of formally defined declarative semantics. The benefit of the knowledge exchange interlingua was that it could allow knowledge from one KBS to be incorporated into another KBS, and it would also allow KBSs to inter-operate so that they could cooperatively perform tasks and solve problems. For example: assertions, queries, and simulation inputs and outputs could all be exchanged.

**L.2 The Knowledge Representation System Specification (KRSS)**

The KRSS working group developed a specification for knowledge representation systems based on description logics. A significant proportion of the collaborators in the KRSS group had experience with descriptive logics (DL) systems and there was already a large body of formal work on DLs that helped to simplify the specification process. Differences noted between the various DL systems reflect different positions taken in the trade-off between expressive power, completeness of inference, and resource consumption. The KRSS negotiated a DL that combined the constructs generally agreed upon by the community. As well they developed a set of interface functions that allowed for the construction, manipulation, and querying of DL knowledge bases.

**L.3 The Knowledge Query and Manipulation Language (KQML)**

The KQML working group designed a common high-level language and associated protocol that could be used by software systems (including intelligent agents and human driven systems) for the run-time sharing of information and knowledge. A special class of agents was used called communication facilitators that amongst other things, maintained a registry of knowledge services offered and sought by different agents.

KQML expressions were layered according to:

- the content i.e. the expression in some agreed Interlingua such as KIF;
- the message i.e. the logic of the communication e.g.
  - to forward a message,
  - to distribute a message,
  - to assert some fact that will cause an agent to perform some reasoning,
o to retract some fact that will cause an agent to perform some reasoning,

o to ask or query an agent,

o to generate answers to some query, or

o to monitor an agent’s goal store and cause the agent to inform the sender when some matching state occurs; and

• the mechanics of the communication i.e. information about the sender, the recipient, a unique ID for the communication, synchronisation information and so on.