

3

Análise das Limitações dos Ambientes de Gestão do Conhecimento

Neste capítulo serão analisadas as principais barreiras e limitações dos sistemas de gerenciamento de conhecimento, como elas podem ser solucionadas através do uso da Web Semântica e como esses problemas e soluções se aplicam no cenário em estudo. Essa análise inicial será descrita na seção 3.2, entretanto, uma breve introdução aos conceitos de Web Semântica será apresentada antes, na seção 3.1.

Para auxiliar na identificação da presença dos fatores limitantes no estudo de caso, uma pesquisa de satisfação foi elaborada e respondida pelos operadores do CORS (usuários do sistema). A pesquisa e os resultados são explanados na seção 3.3.

Ainda na fase de análise, seção 3.4, os principais índices de atendimento e disponibilidade dos serviços de TIC serão levantados e avaliados. Esses índices estão relacionados com o uso do *KMS* em estudo e serão usados para posterior comparação com o novo *KMS* que será desenvolvido baseado nas tecnologias da Web Semântica.

Na seção 3.5 será descrito os requisitos para o novo sistema de gerenciamento de conhecimento baseado em Web Semântica.

3.1.Web Semântica

A maior parte do conteúdo disponível na Web hoje é desenvolvida para leitura e interpretação dos humanos e não para os computadores. Estes estão limitados a identificar e processar determinados elementos e estruturas de documentos, como cabeçalhos e *links*. Através da Web Semântica as máquinas podem entender o significado de um documento e manipulá-lo de forma sensível ao contexto possibilitando a execução de tarefas avançadas para seus usuários. Por exemplo, um agente de software deixa de tratar Brasília, Brasil e capital como

meras *strings* que devem ser exibidas e passa a entender através de formalismos semânticos que Brasília é a capital do Brasil.

Na Web Semântica todos os recursos são representados como meta dados utilizando o modelo *Resource Description Framework* (RDF) [W3C, RDF], que permite que os usuários pesquisem os dados desejados e obtenham os respectivos resultados como se estivessem usando modelos de dados manipulados em SGBDs.

O modelo RDF permite expressar significados através de uma estrutura baseada em triplas compostas por sujeitos, predicados e objetos. A representação RDF usa URIs (*Universal Resource Identifier*) para dar identificações aos sujeitos, predicados e aos objetos (quando não literais). Isto permite que qualquer um possa criar novos conceitos, sendo necessária apenas a definição de uma URI, podendo ou não, estar disponível em algum local na Web. O trecho abaixo é um exemplo de modelo RDF em notação N3 [W3C, N3].

```
@prefix dc: <http://purl.org/dc/elements/1.1/>.

<http://www.monitoracao.org/Matheus_Salcedo >
  dc:title "Matheus Salcedo";
  dc:publisher "Monitoracao".
```

Neste trecho o sujeito “http://www.monitoracao.org/Matheus_Salcedo” tem as propriedades dc:title e dc:publisher cujos respectivos valores são os recursos “Matheus Salcedo” e “Monitoracao”. Note que tanto o sujeito quanto o predicado são identificados por URIs. O prefixo dc: é uma abreviação cuja definição aparece na declaração @prefix, que indica a URI do padrão de meta dados Dublin Core ², isto significa que as propriedades *title* e *publisher* fazem parte de um vocabulário previamente padronizado. Assim, dc:title é uma abreviação para http://purl.org/dc/elements/1.1/title. Este padrão foi criado pela *Dublin Core Metadata Initiative (DCMI)* para descrever objetos digitais. A DCMI é uma organização que busca promover a adoção de padrões de interoperabilidade de meta dados e desenvolver vocabulários para descrever fontes ³.

Uma ontologia define os termos usados para descrever e representar uma área de conhecimento. De acordo com [Grubber, 1993], uma ontologia é uma especificação explícita e formal de uma conceitualização compartilhada. Ainda de acordo com [Grubber, 1993], para a ciência da computação uma ontologia consiste

² e ³ DCMI – www.dublincore.org

de uma hierarquia de conceitos, além de atributos, relacionamentos, restrições, instancias e axiomas que descrevem o conhecimento de um determinado domínio de modo consensual e compartilhado. Esta descrição permite que as máquinas, através de agentes de software, possam entender o significado de um termo (i.e., processá-lo) e estabelecer inferências sobre esses significados. Por exemplo, uma ontologia pode expressar a seguinte regra “o filho de um tio é um primo”. Com esta regra previamente definida um programa pode inferir que se José é meu tio e João é filho de José, então João é meu primo.

As linguagens recomendadas pela W3C para definir e instanciar ontologias são o RDF-Schema (RDFS) [W3C, RDFS] e a Web Ontology Language (OWL) [W3C, OWL]. A primeira é uma extensão do RDF e introduz os conceitos de classes, propriedades, domínios e contra-domínios, entretanto, ainda é limitada para descrever recursos com detalhes. A linguagem OWL possibilita uma maior expressividade semântica na conceitualização de um domínio. O trecho a seguir é um exemplo de regra de inferência desenvolvida em OWL.

```
<owl:ObjectProperty rdf:ID="locatedIn">
  <rdf:type rdf:resource="&owl;TransitiveProperty" />
  <rdfs:domain rdf:resource="&owl;Thing" />
  <rdfs:range rdf:resource="#Region" />
</owl:ObjectProperty>

<Region rdf:ID="Porto Alegre">
  <locatedIn rdf:resource="#Rio Grande do Sul" />
</Region>

<Region rdf:ID="Rio Grande do Sul">
  <locatedIn rdf:resource="#Brasil" />
</Region>
```

O trecho acima ilustra a propriedade transitiva, onde, dado uma função F transitiva, para todo x , y e z : $F(x,y)$ e $F(y,x)$ implica em $F(x,z)$. Neste exemplo a função transitiva “locatedIn” é definida no primeiro parágrafo, em seguida, definimos a região de Porto Alegre como sub-região de Rio Grande do Sul que por sua vez é uma sub-região do Brasil. Portanto, através da propriedade de transitividade é possível inferir que Porto Alegre é uma sub-região do Brasil.

Além deste tipo de propriedade, a linguagem OWL permite especificar outras, tais como as simétricas e inversas e também permite restrições de cardinalidade e disjunção de classes.

3.2.

Fatores Limitantes nos Sistemas de Gerenciamento de Conhecimento

Esta seção fará uma análise sobre as limitações do sistema de gerenciamento de conhecimento introduzidas no capítulo 1. A seção será estruturada em tópicos, cada tópico abordará uma possível limitação dos sistemas de gerenciamento de conhecimento e como a Web Semântica pode solucioná-la, também será identificada a relação desta limitação com o sistema do estudo de caso.

3.2.1.

Pesquisa do Conhecimento

Este fator está relacionado à precisão da busca e a organização do conhecimento procurado. Os mecanismos de buscas sintáticas usados atualmente podem ser ineficientes e não trazer toda a informação desejada relacionada ao item procurado. A organização do conhecimento procurado também pode estar incompleta e a informação desejada ausente.

Através da representação de significado fornecido pelo RDF e RDFS é possível analisar os vocabulários contidos na pesquisa de forma semântica. Isto fornece ao usuário da pesquisa um conhecimento preciso e sensível ao contexto. Diferente dos mecanismos de buscas tradicionais limitados a análise sintática restrita a estrutura das frases e palavras chave. Com auxílio das ontologias, fornecidas pelas ferramentas da Web Semântica, é possível aumentar ainda mais o poder da busca através de regras inferências sobre os conhecimentos. Neste caso, a máquina pode concluir informações mesmo que elas não sejam explícitas. Por exemplo, imagine que o usuário deseja procurar um médico ortopedista especializado em joelhos que atenda no bairro de Copacabana no Rio de Janeiro. Neste exemplo, é difícil encontrar a informação desejada através de uma simples busca sintática. Ao invés disso, o mecanismo de busca pode entender o significado do vocabulário da pesquisa, identificar que médico é um recurso descrito por algumas ontologias, dentre elas uma fornecida pelo Conselho Regional de Medicina. Esta ontologia contém determinadas propriedades como o nome do médico, número do CRM, telefone, e-mail, especialidades, endereço comercial e etc. O próprio Conselho Regional de Medicina pode fornecer uma base de triplas

RDFS que segue esta ontologia e contém os dados dos médicos licenciados. O mecanismo de busca pode então consultar a base de triplas e retornar as informações sobre recurso procurado (médico) que tenha as propriedades (especialidades, endereço comercial) com os objetos desejados (ortopedista, joelhos e Copacabana, Rio de Janeiro). Note que o usuário poderia buscar apenas a especialidade ortopedista e através de uma regra de inferência transitiva o computador poderia inferir que se trata de um médico.

Conforme descrito na seção 2.1, no sistema de gerenciamento de conhecimento em estudo a pesquisa por um procedimento pode ser realizada através de busca sintática ou através de uma navegação no portal. Os conhecimentos estão organizados de maneira hierárquica de acordo com a estrutura organizacional da empresa. A busca é ineficiente e a navegação pela estrutura é lenta e exige conhecimento prévio da estrutura. Como não há relacionamento da base de conhecimento com os outros sistemas, os integrantes do CORS são obrigados a buscar informações em outros sistemas para determinar como devem agir.

Para implementar uma busca semântica e solucionar a limitação da buscas ineficientes do estudo de caso é necessário definir uma ontologia que descreve os recursos envolvidos. A ontologia deve descrever as propriedades dos servidores, conhecimentos, alarmes e o relacionamento entre eles. Também é possível definir algumas regras de inferência. Outra alternativa seria importar e adaptar ontologias de entidades externas como, por exemplo, usar uma ontologia padrão fornecida por determinados fabricantes para descrever servidores.

3.2.2. Integração do Conhecimento

Este fator está relacionado com a combinação e interseção dos dados presentes nas diferentes fontes de dados. O objetivo da integração dos dados é fornecer aos usuários finais um acesso integrado a todos os tipos de dados armazenados em bases de dados heterogêneas. Isto deve ocorrer de forma transparente com uniformidade e integridade, por meio de consultas decompostas e distribuídas, emitindo-as aos componentes responsáveis e retornando os resultados combinados de forma uniforme para os usuários finais.

De acordo com [Giachetti, 2004], existem três níveis possíveis de integração: dados, aplicação e processos. Para garantir a integração no nível de dados é necessário manter um mecanismo de compartilhamento de dados através do qual os diferentes sistemas se comunicam. Para implementar a integração em nível de aplicação é necessário manter a interoperabilidade entre os sistemas. Ainda segundo [Giachetti, 2004], o principal obstáculo para integração surge na heterogeneidade sintática e semântica entre diferentes sistemas e aplicações.

A metodologia tradicional para implantar interoperabilidade entre sistemas heterogêneos no nível de dados, aplicações ou processos é baseada no uso de *middlewares*. Esta metodologia permite uma integração sintática e estrutural em todas as camadas, mas não fornece integração semântica.

O exemplo ilustrado na figura 8 proposto em [J. Joo & S. M. Lee, 2009], compara a metodologia de integração tradicional com a metodologia de integração baseada em Web Semântica. O *middleware* tradicional precisa atuar como um tradutor das diferentes estruturas presentes em cada um dos sistemas para garantir a comunicação entre eles. No pior caso, para n sistemas, podem ser necessárias $n*(n-1)$ operações de mapeamento e tradução de dados quando uma mensagem é enviada de um sistema para os demais. Nesta abordagem o sistema não interpreta o significado dos termos.

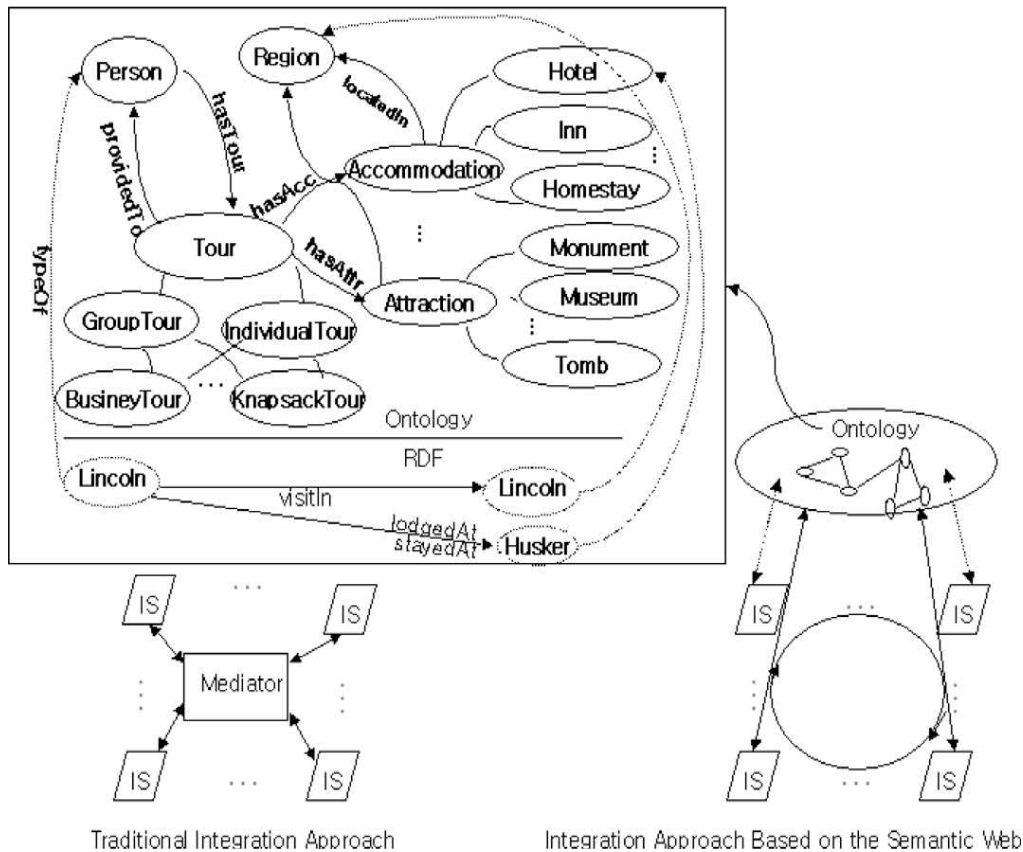


Figura 8 – Metodologia de integração tradicional x Semântica [J. Joo & S. M. Lee, 2009].

Na metodologia baseada em Web Semântica os agentes de software entendem o significado dos dados através do processamento automático a partir do grafo RDF e de ontologias.

Ainda no exemplo da figura 8 de [J. Joo & S. M. Lee, 2009], é usado uma sentença com termos homófonos e homônimos: *President Lincoln visited Lincoln and stayed (or lodged) at The Husker*. Em um determinado sistema *Lincoln* representa o nome de uma pessoa, enquanto em outro sistema *Lincoln* representa o nome de um lugar. Os termos *stayed* e *lodged* são usados por sistemas diferentes e expressam o mesmo significado. O sistema baseado em Web Semântica interpreta o significado dos termos de acordo com as ontologias. Neste exemplo o primeiro *Lincoln* é do tipo *Person* e o segundo do tipo *Region*. *The Husker* além de ser do tipo hotel é uma subclasse de *Acommodation*. *Stayed* e *lodged* significam a mesma coisa.

O estudo de caso nesta tese possui pelo menos três sistemas diferentes (TEC, ITCM e o próprio sistema de gerenciamento de conhecimento) usados para apoiar

o CORS. As bases alimentadas por esses sistemas também armazenam seus dados em estruturas com esquemas distintos, portanto, este cenário é caracterizado por uma heterogeneidade sistêmica e estrutural. Outro possível problema para integrar os dados do ambiente em estudo é a heterogeneidade Semântica onde um mesmo dado pode possuir interpretações diferentes, o que acontece principalmente quando temos palavras homófonas ou homônimas.

Atualmente, não existe nenhuma integração entre o sistema gerenciador de conhecimento de alarmes, a TEC, o ITCM e o Formula. O operador perde um tempo significativo procurando informações ou conhecimentos em todos esses sistemas para tratar os alarmes recebidos aumentando, conseqüentemente, o tempo de indisponibilidade de um serviço.

3.2.3. Tempo, Espaço e Inconveniência

Esses fatores estão relacionados com a disponibilidade do sistema aos locais de acesso e as facilidades de uso. A Web Semântica pode ser combinada com a computação ubíqua para disponibilizar acesso aos sistemas de gerenciamento de conhecimento a qualquer momento e em qualquer local. A computação ubíqua tem como objetivo tornar a interação pessoa/máquina invisível, ou seja, integrar a informática com as ações ou comportamentos naturais das pessoas. A integração da computação ubíqua com a Web Semântica permite uma computação sensível ao contexto. Por exemplo, ao inserir um congelado no microondas, este pode ler um código de barras do produto que contem sua identificação e de seu fabricante (URIs). Em seguida, o microondas pode fazer uma consulta na base de triplas RDF do fabricante para identificar parâmetros ideais para cozinhar o produto, como tempo de preparo e potência.

O acesso ao sistema de gerenciamento de conhecimento atual pode ser feito através de dispositivos móveis, posto que é uma aplicação Web e como outra qualquer, pode ser exibida em diferentes browsers. Entretanto, este acesso não é possível para TEC e ITCM. A navegação através de dispositivos móveis no KMS em estudo é complexa e não possui uma estrutura específica para este tipo de acesso e não fornece todas as informações desejadas.

Através da integração e pesquisa baseada em Web Semântica, o usuário se limitará em informar o nome do servidor, tipo de alarme, item de configuração ou qualquer outro recurso para o sistema. O resultado da pesquisa retornará todo conhecimento existente relacionado ao recurso consultado para o usuário e todo processamento do sistema será transparente. O acesso pode ser feito de qualquer local com acesso à rede da empresa, inclusive através de dispositivos móveis.

3.2.4. Qualidade do Conhecimento

A qualidade do conhecimento é um fator relacionado à relevância, redundância, completude, abstração, validade e confiança do conhecimento. Os sistemas de gerenciamento de conhecimento baseados em Web Semântica podem ser beneficiados com o compartilhamento de conhecimentos entre comunidades diferentes.

A interoperabilidade fornecida pela Web Semântica permite o envolvimento, uso e manipulação intuitiva de estruturas de conhecimento por diferentes comunidades [Neumann, Hogan, & MacDonaill, 2005]. A base de conhecimento usada para tratar alarmes de servidores de uma determinada empresa pode ser enriquecida com o conhecimento de problemas de servidores de seus respectivos fabricantes.

Nonaka & Konno, em 1998, sugeriram o modelo SECI – *Socialization, Externalization, Combination and Internalization* para a criação de conhecimento. A Web Semântica pode ser usada para auxiliar o SECI, conforme a seguir exposto:

- *Socialization* - envolve o compartilhamento do conhecimento entre indivíduos e comunidades. A Web Semântica fornece ferramentas para auxiliar os sistemas de gerenciamento de conhecimento neste sentido, conforme já discutido neste estudo.
- *Externalization* - está relacionado com conversão de um conjunto de informação em conhecimento explícito. As ferramentas disponibilizadas pela Web Semântica (RDF, RDFS e OWL,) possibilitam a representação e a inferência de conhecimentos. Isto facilita o processo de criação de conhecimento.

- *Combination* - envolve a transformação do conhecimento explícito recém-capturado em um conjunto mais complexo, relevante e preciso de conhecimento. Através de mecanismos de integração semântica, descritos na seção 3.2.2, podemos combinar os conhecimentos especializados de diferentes comunidades e indivíduos distribuídos na Web.
- *Internalization* - é o aprendizado, onde o conhecimento armazenado é utilizado pelas pessoas ou máquinas. A Web Semântica permite aumentar a precisão das buscas (seção 3.2.1), isto auxilia na identificação de conhecimentos relevantes ao contexto necessários para o processo de aprendizagem.

A qualidade do conhecimento do *KMS* em estudo deve sofrer impactos positivos com a adoção da Web Semântica. A integração do conhecimento dos especialistas com as informações de configuração e alarmes proporcionará conhecimentos precisos e de melhor qualidade, assim como a busca Semântica em relação à busca sintática.

3.3. Pesquisa de Satisfação

Nesta seção, é desenvolvida uma análise estatística qualitativa sobre as limitações que afetam consideravelmente a qualidade e o conhecimento do sistema em estudo através de uma pesquisa de satisfação respondida pelos usuários do sistema (CORS). O objetivo da pesquisa é o de tentar medir quanto cada limitação influencia no desempenho e satisfação do usuário.

A pesquisa foi dividida em cinco grupos, cada grupo com questões relacionadas aos fatores limitantes identificados e estudados na seção 3.2: tempo e espaço, inconveniência, pesquisa do conhecimento, integração do conhecimento e qualidade do conhecimento, este último inclui incongruência e confiança. No total foram formuladas 21 (vinte e uma) questões. A pesquisa respondida pelos operadores é apresentada na seção 3.3.1, a seguir. Os resultados serão analisados na seção 3.3.2.

3.3.1. A Pesquisa

Para cada questão, as seguintes alternativas estavam disponíveis para resposta: 1- Discordo Totalmente, 2- Discordo Parcialmente, 3- Não Concordo nem Discordo 4- Concordo Totalmente, 5- Concordo Parcialmente. O KMS em estudo é conhecido como 'Portal OI'. Abaixo segue as questões respondidas pelos operadores.

Formulário:

1. Tempo e Espaço

- 1.1. Eu posso usar o Portal OI somente por um tempo limitado.
- 1.2. Eu posso usar o Portal OI somente em determinados locais.
- 1.3. Eu tenho limitações para acessar o Portal OI através de diferentes pontos de acesso, como dispositivos móveis, PDA e Internet.

2. Inconveniência

- 2.1. Eu acho o Portal OI instável por não funcionar de forma correta ou por apresentar erros de sistema.
- 2.2. Eu acho o tempo de resposta do Portal OI insatisfatório (lento).
- 2.3. Eu acho difícil usar o Portal OI.

3. Pesquisa do Conhecimento

- 3.1. Eu acho difícil pesquisar o conhecimento que preciso no Portal OI.
- 3.2. Eu acho o esquema de classificação de conhecimento mal organizado no Portal OI.
- 3.3. Eu acho que o Portal OI fornece conhecimento irrelevante no resultado das pesquisas.
- 3.4. Eu acho que o Portal OI fornece conhecimento redundante ou desnecessário na pesquisa de conhecimento.
- 3.5. Eu tenho alguma dificuldade em encontrar o conhecimento que preciso.
- 3.6. Eu não consigo encontrar conhecimento que preciso quando eu uso o método de procura por palavra chave usando sinônimos.

4. Integração do Conhecimento

- 4.1. O Portal OI não é integrado com outros sistemas, como base de dados de configuração e base de incidentes.
- 4.2. Eu tenho dificuldade de integrar ou agregar o conhecimento do Portal OI com outros sistemas de informação.
- 4.3. Eu tenho dificuldade em integrar o conhecimento do Portal OI com o conhecimento na Web.

5. Incongruência e Confiança (Qualidade do Conhecimento)

- 5.1. Eu acho que o Portal OI contém conhecimentos irrelevantes para minhas tarefas.
- 5.2. Eu acho o conhecimento fornecido pelo Portal OI redundante.
- 5.3. Eu acho o conhecimento fornecido pelo Portal OI não compreensível e nem completo o bastante.
- 5.4. Eu acho o conhecimento fornecido pelo Portal OI muito abstrato.
- 5.5. Eu acho o conhecimento fornecido pelo Portal OI impreciso.
- 5.6. Eu acho o conhecimento fornecido pelo Portal OI não confiável.

3.3.2. Resultados

A pesquisa foi disponibilizada na Web, na intranet da empresa e foi encaminhada para todos os operadores dos centros de comando de todas as regionais. 21 (vinte e um) operadores responderam a pesquisa.

Cada gráfico, abaixo, representa o resultado da pesquisa por grupo de questões, ou seja, por fator limitante. As alternativas ‘Concordo Totalmente’ e ‘Concordo Parcialmente’ foram combinadas em ‘Concorda’, as alternativas ‘Discordo Totalmente’ e ‘Discordo Parcialmente’ foram combinadas em ‘Discorda’, a alternativa ‘Não Concordo nem Discordo’ e ausências de resposta foram consideradas ‘Outros’.

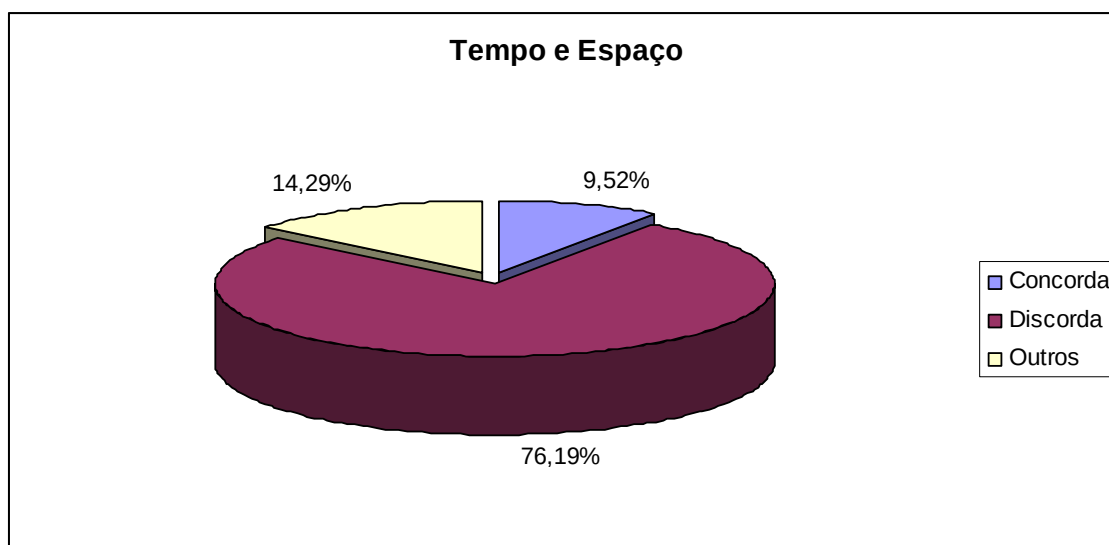


Gráfico 1 – Tempo e Espaço.

O gráfico 1 mostra que 76,19% dos operadores discordam sobre as questões relacionadas à 'Tempo e Espaço' e apenas 9,52% concordam com esta limitação, 14,29% não opinaram ou não concordam nem discordam. De acordo com as respostas dos operadores, o sistema é acessível de locais diferentes e está disponível para consulta quando necessário. Podemos considerar que o resultado indica que o fator limitante 'Tempo e Espaço' não é uma limitação do KMS em estudo.

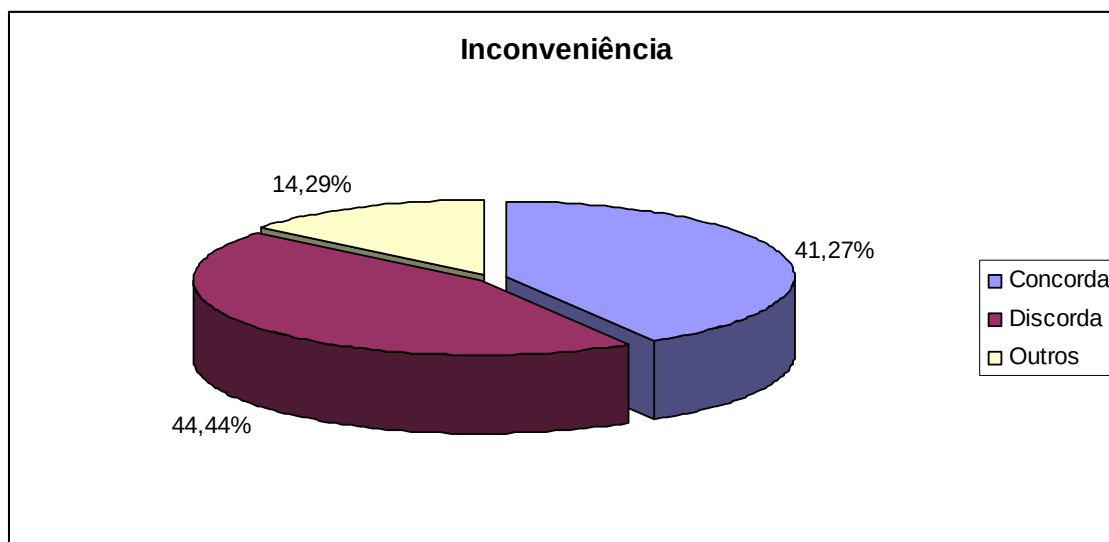


Gráfico 2 – Inconveniência.

O gráfico 2 mostra que 44,44% dos operadores discordam sobre as questões relacionadas à 'Inconveniência', enquanto que 41,27% concordam com esta

limitação e 14,29% não têm opinião formada. Para auxiliar a análise deste tema, segue o resultado das três (3) questões que fazem parte deste grupo:

2.1 – Concorda: 0%, Discorda: 85,71% e Outros: 14,29%.

2.2 – Concorda: 90,48%, Discorda: 0% e Outros: 9,52%.

2.3 – Concorda: 33,33%, Discorda: 47,62% e Outros: 19,05%.

De acordo com as respostas, o *KMS* em estudo é estável e não apresenta falhas, entretanto, o tempo de resposta não é satisfatório e 1/3 dos operadores consideram o sistema de difícil acesso. Podemos concluir que a limitação ‘Inconveniência’ afeta parcialmente o *KMS* em estudo.

O gráfico 3, abaixo, mostra que apenas 3,17% dos operadores discordam sobre as questões relacionadas à ‘Pesquisa do Conhecimento’, a maioria, 90,48%, concordam com esta limitação e 6,35% não tem opinião formada.

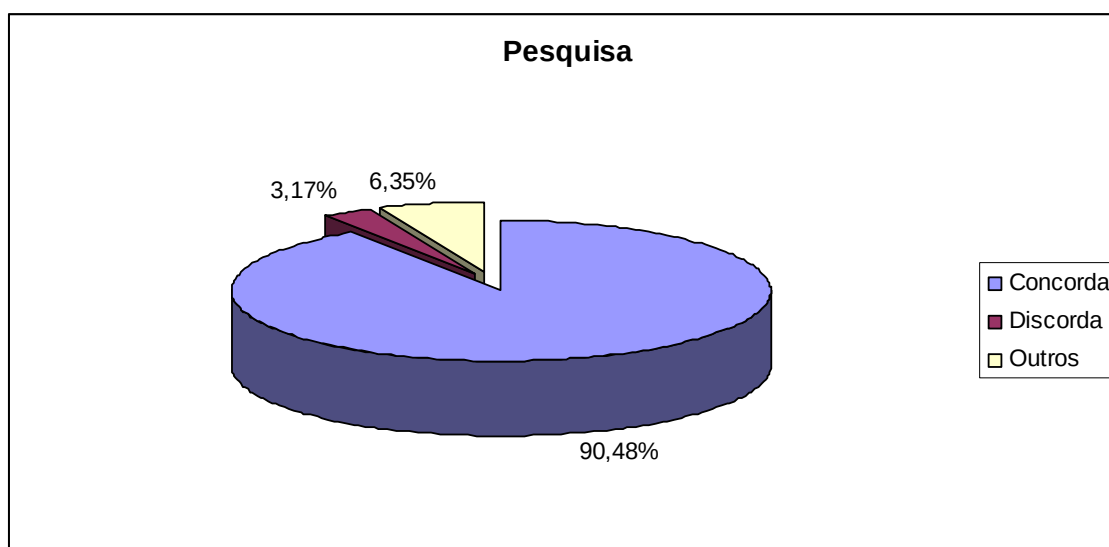


Gráfico 3 – Pesquisa.

De acordo com o resultado da pesquisa, a maioria dos operadores tem alguma dificuldade para encontrar os conhecimentos que procuram. Isto pode ocorrer porque o resultado da pesquisa traz informações desnecessárias, redundantes, irrelevantes, mal organizadas ou existe alguma dificuldade de executar a própria pesquisa. Podemos concluir que as limitações associadas à Pesquisa do Conhecimento afeta o *KMS* em estudo integralmente.

O gráfico 4 mostra que nenhum operador (0%) discorda sobre as questões relacionadas à 'Integração', quase todos, 98,41% concordam integralmente com esta limitação e apenas 1,59% não tem opinião formada (equivalente a uma resposta das 3 questões formuladas para este grupo).

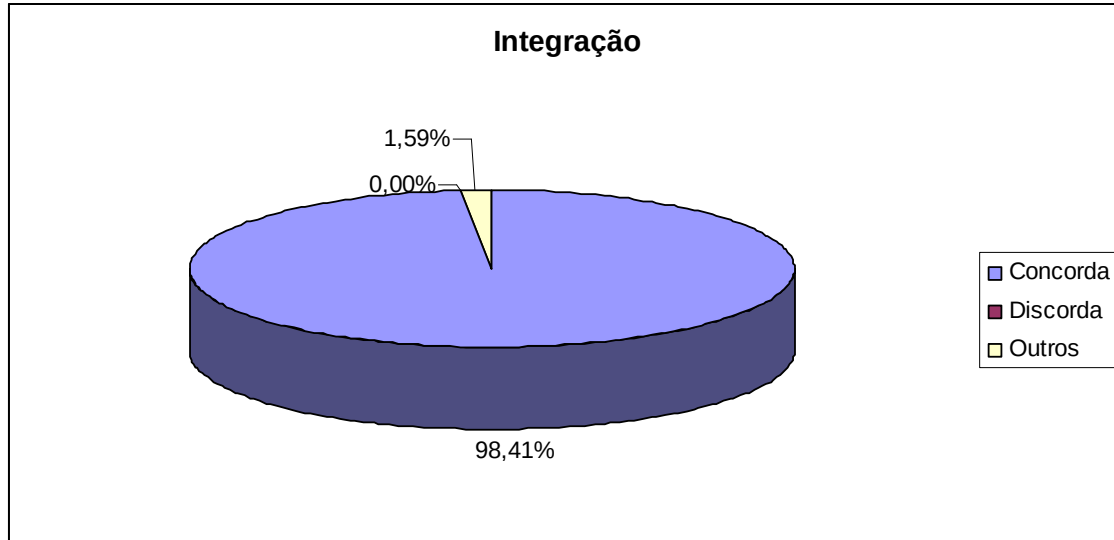


Gráfico 4 – Integração.

De acordo com o resultado da pesquisa, a maioria dos operadores percebe que o *KMS* em estudo não é integrado com outros sistemas de gerenciamento de conhecimento. Em nosso estudo de caso este fator é crítico, a ausência de integração do *KMS* com a base de configuração e com a base de alarmes prejudica diretamente o tratamento de alarmes pelo CORS. O operador precisa interagir com os 3 (três) sistemas de apoio para tratar o alarme quando um único sistema integrado poderia executar toda a tarefa que hoje é feita manualmente. Com este resultado, podemos concluir que as limitações relacionadas ao fator 'Integração' estão presentes no *KMS* em estudo.

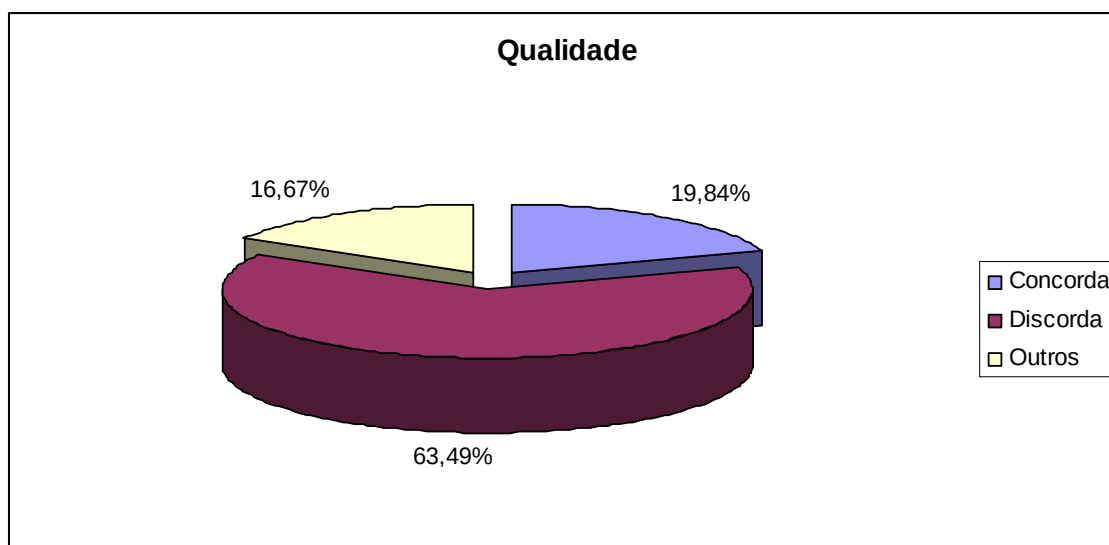


Gráfico 5 – Qualidade.

O gráfico 5 mostra que 63,49% dos operadores discordam sobre as questões relacionadas à 'Qualidade do Conhecimento, 19,84% concordam com esta limitação e 16,67% não tem opinião formada.

O resultado da pesquisa mostra que a maioria dos operadores acham a qualidade do conhecimento do sistema em estudo satisfatória.

3.4. Indicadores

Nesta seção, vamos analisar alguns indicadores que podem ser usados para medir a eficiência do sistema de gerenciamento de conhecimento em estudo. Os indicadores serão úteis para posterior comparação com o novo sistema baseado em Web Semântica. Os seguintes indicadores serão avaliados:

1. Número de ações de primeiro nível tomadas pelo CORS / número de alarmes recebidos.
2. Índices de Disponibilidade.
3. Tempo Médio de Duração de Alarmes.

O primeiro indicador mostra o conhecimento que o operador do CORS tem ou aprende através do sistema de gerenciamento de conhecimento para tomar alguma ação para tratamento do alarme. Essas ações são denominadas de primeiro

nível, pois são executadas diretamente pelos operadores e não são encaminhadas para os especialistas de cada ambiente (pelo menos inicialmente). Ações de primeiro nível diminuem o tempo de indisponibilidade dos serviços de TIC, reduzem o impacto da indisponibilidade para o negócio e reduzem o custo de manutenção da infraestrutura de TIC. Quanto maior este indicador melhor o desempenho do sistema de gerenciamento de conhecimento em estudo que suporta o tratamento dos alarmes.

O segundo indicador reflete os índices de disponibilidade calculados pelo Formula (citado no capítulo 2). Os índices de disponibilidade são calculados mensalmente por servidores e por serviços. A disponibilidade de um servidor é calculada pela seguinte formula:

$$\left(1 - \frac{\sum i}{t}\right) * 100$$

- $\sum i$ = Somatório do tempo de duração de alarmes de indisponibilidade de determinada aplicação de determinado servidor em determinado mês.
- t = tempo total de duração de determinado mês.

A disponibilidade de determinado serviço é um pouco mais complicada de calcular, pois envolve a arquitetura específica do serviço. Por exemplo, um serviço em *cluster* com balanceamento de carga composto por 10 nós, a indisponibilidade de um nó não afeta a disponibilidade do serviço, pois o processamento dos dados é redistribuído para os outros 9 nós.

O segundo indicador, índices de disponibilidade de serviços, deve ser comparado em longo prazo. O índice de disponibilidade pode variar bastante de acordo com as manutenções programadas e problemas de infraestrutura de uma maneira geral. Comparar o índice de disponibilidade de um determinado serviço em determinado mês com o índice do mês seguinte pode não significar alguma mudança no conhecimento do CORS.

O terceiro indicador reflete o tempo médio de reparo de um incidente de monitoração, este contador é iniciado no momento que um alarme é detectado pelo ambiente de monitoração e finalizado quando a monitoração detecta que aquele incidente foi corrigido.

3.4.1. Medição

Ações em Primeiro Nível

Os quantitativos de alarmes recebidos e tratados foram extraídos do ARS. Conforme descrito no capítulo 2, a TEC registra os alarmes como incidentes de monitoração no ARS, este por sua vez, armazena todo histórico de tratamento dos incidentes pelo CORS ou pelos especialistas responsáveis. O termo incidente ou incidente de monitoração será usado para descrever os alarmes registrados no ARS.

O gráfico 6, abaixo, ilustra as estatísticas de atendimento de incidentes pelo CORS em três meses. No mês de Junho de 2010 o CORS recebeu 28.272 incidentes. Desse montante, 7.010 foram tratados pelo CORS com base em algum procedimento registrado no sistema de gerenciamento de conhecimento e 6.206 foram cancelados no ARS pelo CORS.

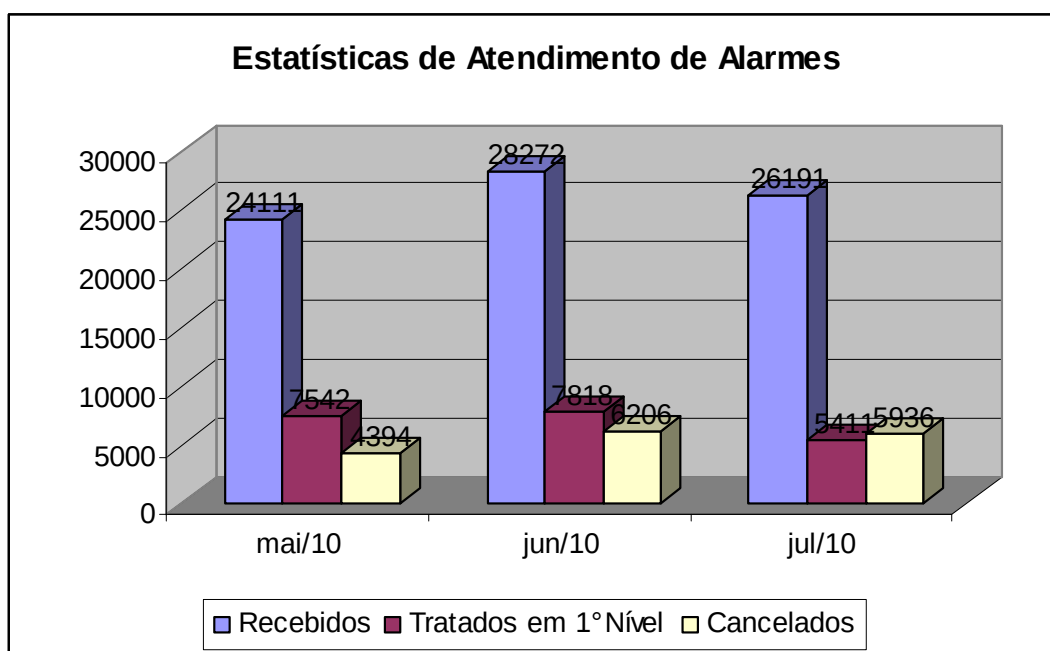


Gráfico 6 – Estatísticas de Atendimento de Alarmes.

Os incidentes são cancelados em algumas situações, dentre elas, quando o alarme é esperado, ou seja, quando um determinado serviço ou servidor está em manutenção programada e a parada do sistema é prevista. O alarme também pode ser cancelado por algumas deficiências nos sistemas de monitoração, por exemplo,

quando os *links* entre determinados CPDs apresentam problemas. O *Nagios* pode reportar milhares de indisponibilidades associados aos servidores instalados no CPD isolado, isto acontece por ainda não existir um mecanismo de correlacionamento entre alarmes de roteadores e alarmes de servidores.

Note que a diferença entre os incidentes recebidos, os incidentes tratados pelo CORS e os incidentes cancelados ($28.272 - 7.010 - 6.206 = 15.056$) são os incidentes repassados pelo CORS para os especialistas sem tratamento prévio. Esses incidentes não possuem nenhum conhecimento associado, encontrado ou interpretado pelo CORS.

O gráfico 7, abaixo, representa o percentual médio referente aos três meses ilustrados no gráfico 6. Dos 78.574 incidentes recebidos nesses três meses, 52,52% foram repassados diretamente pelo CORS para os especialistas sem nenhuma atuação prévia, 26,43% foram tratados em primeiro nível pelo CORS, ou seja, não foi necessário atuação dos especialistas e 21,05% foram cancelados.

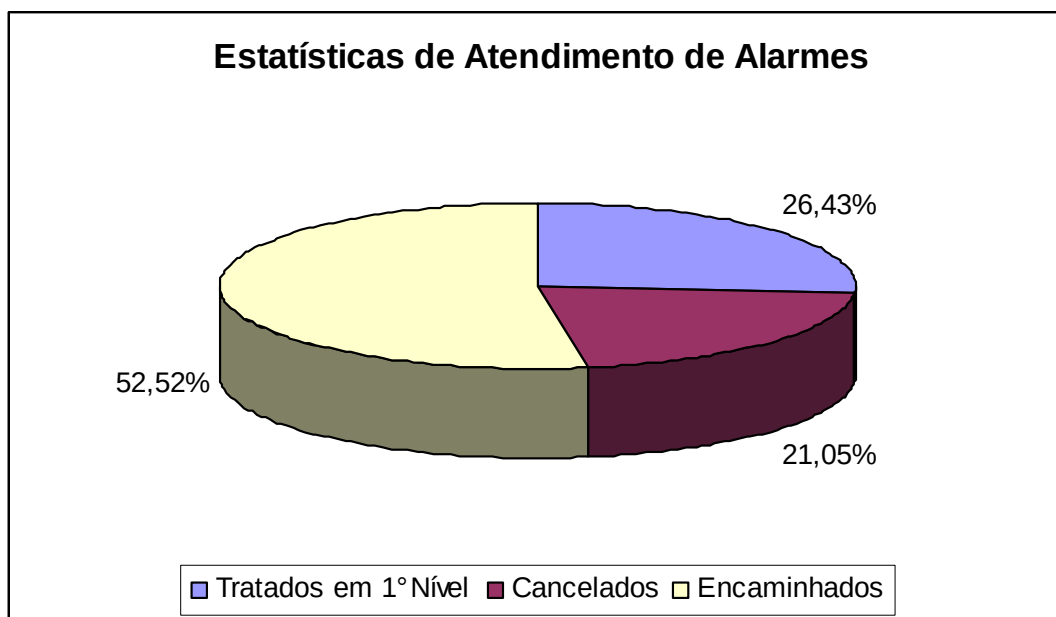


Gráfico 7 – Estatísticas de Atendimento de Alarmes Médio Percentual.

Os dados mostram a ausência de conhecimento do CORS para tratamento dos alarmes, o que possivelmente está relacionado à deficiência do sistema de gerenciamento de conhecimento. A falta de conhecimento do CORS aumenta o custo de manutenção dos ambientes, pois é necessária uma quantidade de homens-

hora maior com o envolvimento dos especialistas, também aumenta o impacto para o negócio, pois aumenta o tempo de indisponibilidade dos serviços.

Para facilitar a posterior avaliação do novo sistema de gerenciamento de conhecimento, baseado em formalismos da Web Semântica, vamos ilustrar as estatísticas de atendimento de alarmes do estudo de caso. Os gráficos 8 e 9, ilustram as estatísticas de atendimento de alarmes dos servidores e serviços do próprio ambiente de Monitoração, descrito no capítulo 2.

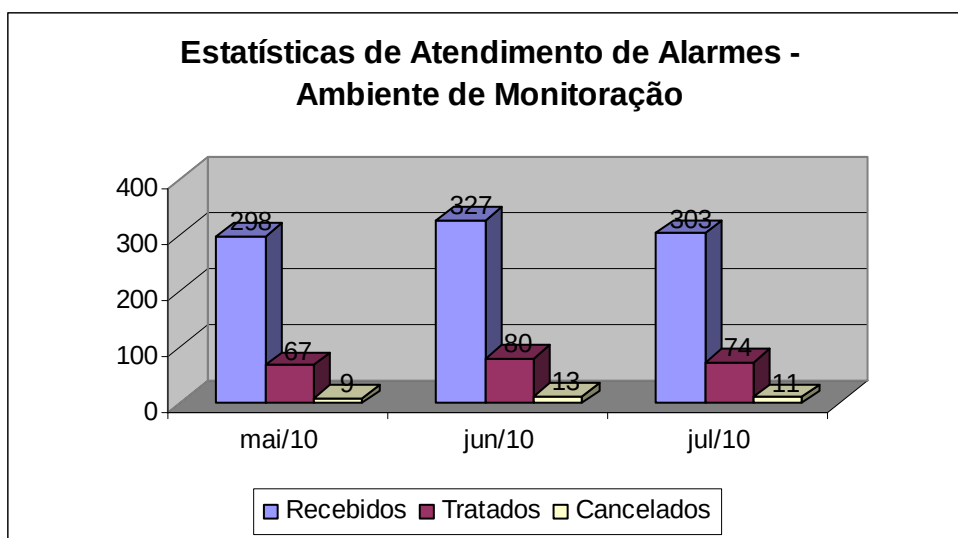


Gráfico 8 – Estatísticas de Atendimento de Alarmes – Ambiente de Monitoração

De acordo com o gráfico 8, no mês de Junho de 2010 o CORS recebeu 327 incidentes do próprio ambiente de monitoração. Note que isto representa pouco mais de 1% do volume total de alarmes. Desse montante, apenas 80 foram tratados pelo CORS com base em algum procedimento registrado no sistema de gerenciamento de conhecimento. Apenas 13 foram cancelados no ARS pelo CORS.

O gráfico 9, abaixo, representa o percentual médio referente aos três meses ilustrados no gráfico 8. Neste período foram registrados 928 incidentes do ambiente de monitoração, 72,63% foram repassados diretamente pelo CORS para os especialistas sem nenhuma atuação prévia, 23,81% foram tratados em primeiro nível pelo CORS e 3,56% foram cancelados.

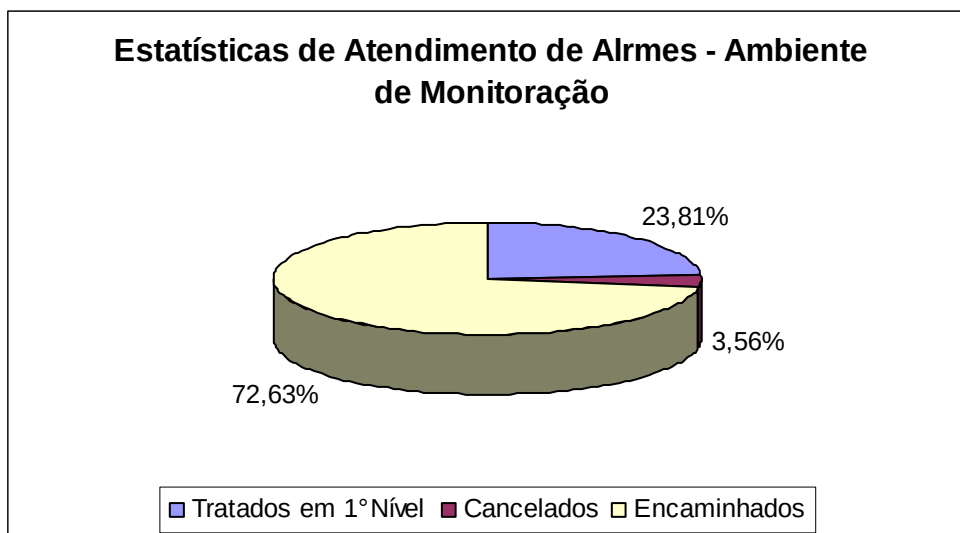


Gráfico 9 – Estatísticas de Atendimento de Alarmes Médio Percentual – Ambiente de Monitoração.

Note que o percentual de incidentes encaminhados sem tratamento prévio no ambiente de monitoração é muito maior que o percentual total ilustrado no gráfico 7, houve um acréscimo de quase 20%.

Disponibilidade de Serviços

Muitos serviços têm índices de disponibilidade medidos e acompanhados, normalmente esses serviços estão atrelados a acordos de níveis de serviço entre a TI e o órgão cliente. Para efeito de estudo de caso, vamos usar o índice de disponibilidade do ambiente de Monitoração, descrito de forma resumida no capítulo 2.

O ambiente de Monitoração é formado por cerca de 250 servidores espalhados por todo Brasil e alguns países do exterior. Conforme descrito no capítulo anterior, o ambiente é formado pelos seguintes sistemas: ITM, ITCAM, Nagios, Vantage e Patrol (o ITCM não faz parte da monitoração).

A tabela 1, a seguir, ilustra os índices de disponibilidade do serviço Monitoração nos últimos meses.

	mai/10	jun/10	jul/10
Disponibilidade do Serviço - Monitoração	99.863%	98.427%	99.674%

Tabela 1 – Disponibilidade do Serviço de Monitoração.

Em Maio/2010 o índice de disponibilidade do sistema foi de 99,863%, isso equivale a pouco mais de uma hora de indisponibilidade. Em Junho, o ambiente marcou 98,427% de disponibilidade, equivalente a pouco mais de 11 horas de indisponibilidade. Em Julho, o ambiente teve um índice de disponibilidade de 99,674%, equivalente a 2 horas e 20 minutos de indisponibilidade no mês.

Tempo Médio de Duração de Alarmes

A tabela 2, a seguir, mostra as estatísticas relacionadas aos tempos de duração dos alarmes dos últimos seis meses, retirados da TEC. O tempo médio de duração de um alarme é de 5 horas, a mediana de 5,02 minutos (0,08 horas) e o desvio padrão é de quase 51 horas.

	Segundos	Minutos	Horas
Média	18014,00	300,23	5,00
Mediana	301,00	5,02	0,08
Desvio Padrão	183556,00	3059,27	50,99

Tabela 2 – Estatísticas de Duração de Alarmes.

O alto desvio padrão mostra uma grande dispersão entre os tempos de duração dos alarmes. Pelo desvio padrão e mediana, podemos interpretar que a média é altamente influenciada pelos altos tempos de duração de determinados alarmes, provavelmente, não críticos. O maior objetivo de um novo sistema de gerenciamento é o de diminuir o tempo de duração de alarmes críticos através do provimento de conhecimento ao CORS. Infelizmente não é possível fazer a distinção dos alarmes críticos e não críticos nas estatísticas acima. Entretanto,

podemos usar a mediana como base de comparação entre o sistema atual e novo sistema.

As estatísticas de duração de alarmes são extraídas diretamente da TEC e não do ARS. Ou seja, estamos avaliando a duração de eventos de monitoração e não de incidentes de monitoração. Muitos eventos não geram incidentes. Enquanto o CORS trata cerca de 25 mil incidentes de monitoração por mês (como descrito acima), a TEC trata cerca de 2 milhões de eventos por mês.

3.5. Requisitos

Um novo sistema baseado em Web Semântica será desenvolvido ou adaptado para substituir o sistema de gerenciamento de conhecimento do estudo de caso. Nesta seção serão descritos os requisitos para este novo sistema.

De acordo com a pesquisa de satisfação realizada no item 3.3.1 as principais limitações presentes no sistema de gerenciamento de conhecimento em estudo estão associadas à Pesquisa do Conhecimento e à Integração, ambas tiveram uma rejeição maior que 90%. A principal meta do novo sistema baseado em Web Semântica será o de aumentar a satisfação do usuário melhorando esses dois principais fatores.

De acordo com nossa análise sobre o fator Integração (item 3.2.1), o sistema atual não é integrado com o ITCM (configuração) e a TEC (alarmes), sistemas que possuem informações essenciais para o tratamento dos alarmes. A ausência de integração desses sistemas afeta diretamente o tempo de duração dos alarmes e a disponibilidade dos ambientes. Para solucionar esta limitação será necessário implementar um *middleware* semântico semelhante ao da figura 4, esta camada intermediária deverá integrar os dados da camada de dados e funcionar como um ponto central para a camada de aplicação que fará interface com o CORS. Essa abordagem será mais bem detalhada no capítulo 5.

Para aumentar a eficiência e a qualidade da pesquisa do conhecimento, além de implementar a integração dos sistemas, os dados das bases de conhecimento, alarmes e configuração serão traduzidos para triplas RDFS, isto significa que os dados poderão ser tratados semanticamente e manipulados através de uma ontologia (o item 3.2.2 abordou bem esta limitação).