

## PROPOSTA ARQUITETURAL DE SISTEMA INTELIGENTE DE ACONSELHAMENTO PEDAGÓGICO APLICANDO RACIOCÍNIO BASEADO EM CASOS

**E. Moura<sup>1</sup>, J. A. da Cunha<sup>2</sup> e C. Analide<sup>3</sup>**

<sup>1</sup>Diretoria Acadêmica de Gestão e Informática – Instituto Federal do Rio Grande do Norte

<sup>2</sup>Mestre em Engenharia de Produção – Universidade Federal de Santa Catarina

<sup>3</sup>Doutoramento em Informática – Universidade do Minho

eli@embits.com.br

Artigo submetido em julho/2014 e aceito em xxxx/2014

### RESUMO

Este trabalho apresenta proposta de aplicação de Raciocínio Baseado em Casos (em inglês, *Case-Based Reasoning*), afim de buscar soluções consolidadas, para aconselhamento pedagógico, dos casos acadêmicos apresentados, facilitando substancialmente a função de acompanhamento e aconselhamento da equipe pedagógica. Pretendemos através desse trabalho

apresentar as motivações que nos levaram a decidir por esta estrutura de algoritmos e dados, de forma a justificar um sistema inteligente e incremental que vai aprendendo as melhores soluções para novos casos a medida que é utilizado, demonstrando as técnicas e tecnologias que permitiram seu desenvolvimento

**PALAVRAS-CHAVE:** Raciocínio Baseado em Casos, Educação, Inteligência Artificial, Aplicativo, CBR.

### ARCHITECTURAL PROPOSAL OF INTELLIGENT SYSTEM FOR EDUCATIONAL ADVISING USING CASE-BASED REASONING ALGORITHM

### ABSTRACT

This work introduces a proposal scheme using Case-Based Reasoning, to find consolidated solutions before used for the new problems for now and future, making easier the task of advising students to the pedagogical staff. We do intend, through this work,

introduce the motivation behind the choices for this system structure, justifying an incremental and smart system who learns bests solutions for new cases when it's used, showing technics and technology.

**KEYWORDS:** Case-based Reasoning, Education, Artificial Intelligence, Machine learning.

## PROPOSTA ARQUITETURAL DE SISTEMA INTELIGENTE DE ACONSELHAMENTO PEDAGÓGICO APLICANDO RACIOCÍNIO BASEADO EM CASOS

### INTRODUÇÃO

Tom Mitchel define aprendizado de máquina como um sistema inteligente onde o sistema (ou programa) aprende da experiência **E** em relação a classe de tarefas **T** e performance **P**, se essa performance nas tarefas em **T**, como medidas em **P**, melhoram com a experiência **E** (Nagy *et al* 2013).

Aprendizado de máquina pode ser aprendizado supervisionado, onde a saída correta no conjunto é colocada à disposição. O outro tipo é o aprendizado não-supervisionado onde a saída exata é desconhecida.

Aprendizado supervisionado é utilizado para resolver classificação de casos ou regressões, como reconhecer spams ou reconhecimento de impressões digitais. Uma das técnicas usadas para aprendizado é um algoritmo de aprendizado supervisionado para geração de uma árvore de decisão. Segundo Nagy *et al* (2013), o algoritmo de aprendizado com árvore de decisão foi proposto primeiramente por Ross Quinlan como uma extensão do algoritmo ID3. Pode lidar com atributos contínuos e discretos, além de conjunto de dados com valores de atributos desconhecidos.

Aprendizado não-supervisionado é utilizado para resolver casos agregados. Uma técnica amplamente utilizada é um algoritmo conhecido como *K-means Clustering*. Esse algoritmo é baseado na distância que separa **n** conjuntos de dados em um número de clusters (agrupamentos) específicos que por sua vez são associados a um determinado centroide e cada grupo de dados segue o agrupamento do centroide mais próximo. O algoritmo *k-means clustering* pode ser descrito da seguinte maneira:

- a) Selecione **K** pontos como o repetidor centroide inicial
- b) Forme **K** clusters por pontos de similaridade com o centroide mais próximo
- c) Recalcule o centroide de cada cluster até que os centroides não mudem mais

O algoritmo *k-means* apresenta-se como a solução que melhor se adapta para o objetivo de indexação, recuperação, revisão e aprendizado de casos por similaridade, proporcionando uma coesão natural com as fases do processo de raciocínio baseado em casos.

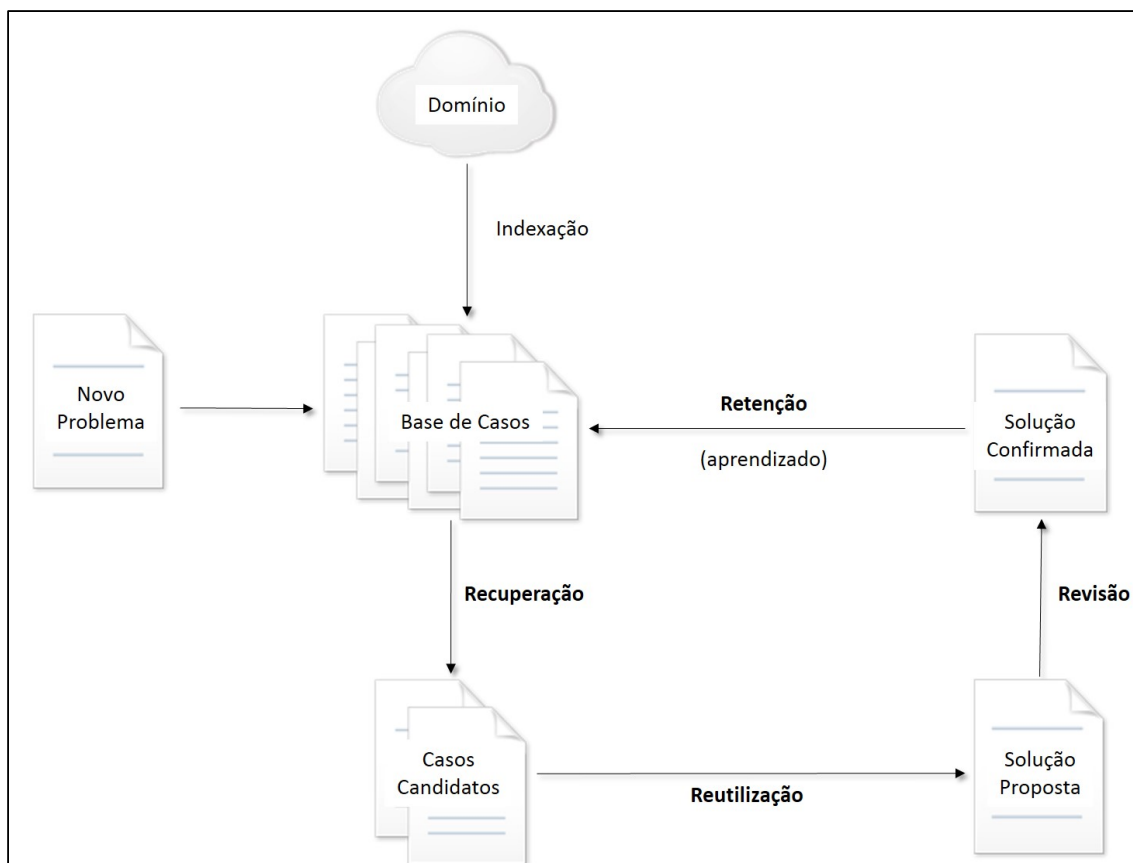
### RACIOCÍNIO BASEADO EM CASOS

Sistemas baseados em conhecimento são sistemas de inteligência artificial que raciocinam sobre suas possíveis ações no mundo fazendo uso de dois mecanismos principais: 1) Base de conhecimento e 2) Mecanismo de inferência.

As principais diferenças entre estes e os sistemas convencionais está na organização dos dados, na separação do conhecimento e método de solução. Sistemas baseados em

conhecimento utilizam soluções de raciocínio que podem ser baseados em regras, em modelos, em casos etc.

Raciocínio Baseado em Casos resolve casos ao recuperar e adaptar experiências passadas – chamadas casos – armazenadas em uma base de casos. Um novo caso é resolvido com base na adaptação de solução de casos similares já conhecidos (Wangenheim *et al* 2013).



**Figura 1 – Ciclo do Raciocínio Baseado em Casos**

Segundo Watson (Watson 1997), Raciocínio Baseado em Casos pode ser resumido como um processo cíclico que compreende os quatro R's, conforme descrito na Figura 1:

- a) Recuperação dos casos mais similares
- b) Reutilização do caso para solucionar o caso
- c) Revisão da solução proposta
- d) Retenção (aprendizado) da nova solução como parte de um novo caso.

Para tanto é necessário que exista uma base de casos e um mecanismo de inferência que através de similaridade apresente soluções para o caso apresentado fazendo uso, quando necessário, de adaptação dos casos existentes ao dado caso.

## Caso

Caso é uma unidade contextualizada de conhecimento que representa experiência (Watson 1997), e deve conter: o caso que descreve o estado das coisas quando ocorreu o caso e a solução que foi estabelecida para o caso. Um caso diferencia-se das outras experiências armazenadas contribuindo com uma melhor solução para o caso apresentado; um caso não é genérico, pois suas informações são contextualizadas a uma determinada situação-caso; um caso é uma instância concreta de um caso.

Como primeiro passo na construção da base de casos, está a aquisição e representação dos casos. No quesito aquisição, podemos encontrar diversos cenários que podem incluir: 1) Os casos não estão disponíveis em uma fonte externa; 2) Os casos estão disponíveis parcialmente em uma fonte externa; 3) Os casos estão disponíveis e contém erros; 4) Os casos estão disponíveis e estão corretos. Depois de coletados os casos, deve-se definir a forma como serão representados.

Não existe um consenso a respeito de que tipo de dados ou quais informações devem estar contidas em um caso, mas devemos ter em mente dois aspectos significativos destas informações: a facilidade de aquisição da informação e a funcionalidade desta.

Um caso ideal pode ser estruturado da seguinte forma: Uma descrição dos aspectos relevantes do caso; o contexto no qual o caso está inserido; a descrição da solução associada ao caso e a avaliação da solução empregada.

## Base de casos

A estrutura de dados composta de casos é chamada Base de Casos, pode ser uma base de dados em forma de documentos, pares de chave-valor, objetos, predicados, redes semânticas, etc. O mais utilizado é o caso representado como um objeto, ou um frame, embora seja comum combinar um ou mais formalismos entre os citados.

Na estrutura de dados orientado a documentos, uma notação usada amplamente e que será usada na solução proposta é a JSON (*Javascript Object Notation* – Notação de Objetos Javascript), que permite descrição de objetos como documentos, agrupados em conjuntos de chave-valor, onde o valor pode ser um documento ou objeto agregado e subsequentemente.

## Indexação

A indexação é um processo importante e decisivo dentro do RBC, assim como numa base de dados convencional, porém não necessariamente, todas as informações relevantes deverão ser indexadas. As informações indexadas servirão para acelerar o processo de recuperação enquanto que informações não-indexadas podem servir de contextualização para a decisão de reutilização do caso ou em outros aspectos do sistema de informação e não somente na

recuperação.

Como identificar quais são as informações que deverão indexar os casos? Não existe uma resposta simples de se obter, porém devemos nos ater a algumas características que índices devem possuir: ser preditivo; estar direcionado ao caso ao qual será aplicado; ser abstrato o suficiente para que através de extrapolação possa ser usado; ser concreto o suficiente para ser reconhecido no futuro (Wangenheim *et al* 2013). Devemos ter em mente que os casos serão indexados de forma automatizada e que existem cada dia mais métodos de indexação, por medidas ou especificações, baseadas em diferenças, por similaridade e métodos indutivos de aprendizado são alguns exemplos.

### **Recuperação**

A recuperação deve utilizar um algoritmo que verifique os clusters que tem maior similaridade com o caso-caso apresentado e permita otimizar a reutilização e revisão do caso. Essa recuperação pode se dar como uma busca por proximidade, busca por simulação paralela, busca hierárquica, busca serial etc. Ao contrário de um banco de dados relacional, uma base de casos requer que a recuperação dos casos se dê de forma avaliada, envolvendo heurísticas e julgamentos sobre a relevância do dado recuperado. Entre os algoritmos mais relevantes e divulgados na recuperação de casos estão: Recuperação de Padrões, Algoritmo de Indução e Algoritmo de vizinhança.

### **Reutilização**

Raciocínio baseado em casos aplica-se na solução de casos pela possibilidade de que, caso haja conhecimento do domínio aplicado ou das constantes existentes neste, pode-se aplicar uma técnica interpretativa através da qual podemos reutilizar uma solução conhecida de um caso semelhante, ou adaptar uma ou mais soluções conhecidas para propor a solução para o caso apresentado. Existem duas etapas no processo de reutilização: Adaptação e Justificação e Crítica (Abel e Castilho 1996).

### **Revisão**

Revisão é o processo de julgar quão boa uma solução é. Neste passo a solução deverá ter sido experimentada no mundo real e recebe um feedback sobre o que ocorreu e se foi correspondente ao esperado, ou se gerou alguma situação nova como resultado. Neste caso, será de suma importância receber o feedback sobre a anomalia com descrição detalhada dos fatores e como prevenir o surgimento de novas anomalias. Essas informações podem ser: explicação das diferenças, justificativas diferentes, comparação de possibilidades alternativas e avaliação

### **Retenção**

## PROPOSTA ARQUITETURAL DE SISTEMA

Reconhecer as limitações do seu tempo é um importante vetor para a compreensão dos limites da ciência e de como os resultados obtidos podem ser influenciados pelas técnicas e tecnologias aplicadas ao método. Os dias atuais trazem consigo tecnologias de armazenamento, processamento e distribuição de dados através de diversas plataformas e com uma velocidade nunca antes vista.

### Linguagem de Programação

Propomos utilizar uma linguagem de programação de ampla aceitação, que permita a implementação independentemente da plataforma utilizada, e que seja de fácil uso e compreensão é um ponto a ser levado em consideração. Dentre as mais conhecidas, é lugar-comum a opinião de que a linguagem Javascript, na sua especificação mais recente, é linguagem de programação poderosa, suficientemente complexa e que permite sua aplicação tanto compilada em servidores e máquinas virtuais, quanto como aplicações clientes, sendo interpretadas em navegadores de internet comuns.

Dentre as opções possíveis, o Javascript é a linguagem candidata perfeita, visto que pode ser aplicada em navegadores de internet, em sistemas operacionais de computadores, *smartphones* e *tablets*.

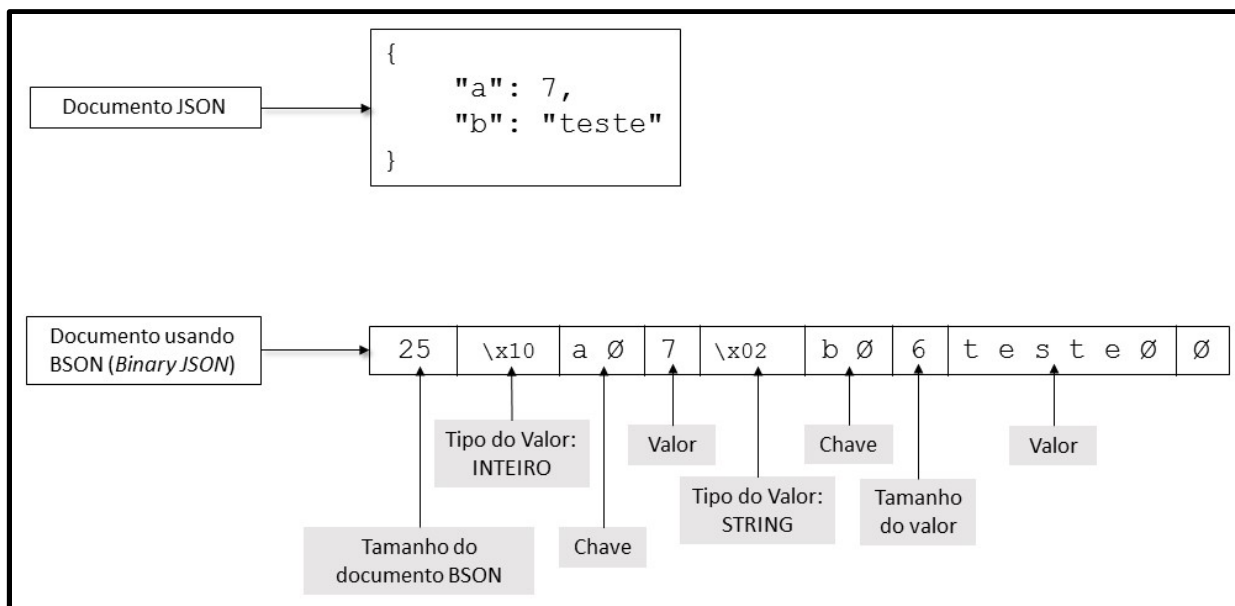
### Persistência de Dados

Como um fator importante ao RBC, a base de casos sendo persistida em um banco de dados que mantém sua estrutura, como coleção de documentos, e ainda permitir a utilização nativa de algoritmos de classificação, mapeamento e redução, nos auxilia a obter resultados que anos antes, com uso de outras tecnologias seriam mais difíceis e mais custosas de se obter.

Propomos a utilização de uma base de dados noSQL orientada a documentos como mecanismo de persistência dos casos utilizados na solução proposta. O termo noSQL é usado amplamente para descrever estruturas de bancos de dados que vão além da estrutura relacional, e que podem incluir conceitos de tuplas, grafos, pares de chave-valor, árvore, documento etc.

Dentre as soluções disponíveis no mercado, a solução que mais se assemelha ao proposto é um SGBD (Sistema de Gerenciamento de Banco de Dados) não-relacional, orientado a documentos, de código aberto e uso livre. O SGBD MongoDB fornece nativamente métodos de mapeamento e redução, que permite obter resultados otimizados e maior velocidade de processamento com redução de recursos.

O MongoDB utiliza também o Javascript como linguagem nativa para suas rotinas e aplicações, além de armazenar os dados em um formato semelhante à notação de objetos Javascript (JSON) só que em binário ao invés de texto pleno. A Figura 2 ilustra a comparação entre um documento representado em texto pleno e sua representação em binário utilizando a notação Javascript.



**Figura 2 – Comparação da representação de um mesmo documento usando JSON em texto pleno e em binário.**

### Classificação

Como metodologia de facilitação para indexação e recuperação dos casos, adotamos a classificação dos casos dividindo-os em agrupamentos chamados classes, levando em consideração às demandas relacionadas a cada caso e determinando um peso a cada demanda. Com isso poderemos determinar um valor para cada classe, de acordo com a quantidade de demandas e o valor individual das demandas de cada classe.

Seguindo o conceito dos quatro elementos de um caso ideal: a descrição dos aspectos relevantes do caso; o contexto no qual o caso está inserido; a descrição da solução associada ao caso e a avaliação da solução empregada, definimos os casos dentro do domínio de aconselhamento educacional conforme os termos empregados pelas especialistas consultoras deste trabalho.

Dessa forma, o contexto pode ser considerado como uma ou mais classes de demandas, onde cada demanda descreve um aspecto do caso analisado. As soluções propostas serão compostas por um ou mais encaminhamentos, enquanto que a classificação de cada caso é

composta pelo peso da classe do caso multiplicado pela soma de pesos de cada demanda que compõe o caso.

A classificação em classes de demandas foi realizada manualmente, por um especialista da área do domínio da aplicação. As demandas iniciais estão listadas em ordem alfabética na Lista 1 como segue:

- Atrasos constantes
- *Bullying*
- Conflito com relação a opção sexual
- Conflito familiar
- Conflito relacionado a situação afetiva, sexual ou de relacionamento
- Desequilíbrio ou caso de ordem psicológica
- Desmotivação pela opção de curso
- Desmotivação por baixo rendimento
- Dificuldade de aprendizagem em disciplina
- Muitas faltas
- Necessidade de orientação pedagógica
- Necessidade de orientação secular (pessoal, não relacionado com a vida acadêmica)
- Pais em separação
- Caso de comportamento
- Caso de ordem disciplinar grave
- Caso de ordem disciplinar leve
- Caso de ordem disciplinar médio
- Caso de relacionamento aluno x aluno
- Caso de relacionamento com a mãe
- Caso de relacionamento com o pai
- Caso de relacionamento em casa
- Caso de relacionamento professor x aluno
- Caso de ordem socioeconômica do aluno ou da família
- Situação de abuso (moral, sexual etc.)
- Situação de exclusão em sala
- Situação relacionada à timidez

#### **Lista 1 – Demandas em ordem alfabética**

As classes resultantes desta classificação são distintas na natureza de suas demandas e também na quantidade de demandas que compõe cada classe. Também foi atribuído um valor (peso) para cada demanda, baseado em seu nível de necessidade de atenção, levando ao prognóstico de que demandas com valores mais altos dentro de cada classe carecem de atenção maior por parte dos especialistas do que as que tem valor menor.

Da mesma forma, as classes de valores mais altos necessitam de maior atenção. Com este mecanismo de classificação simples, conseguimos distinguir casos menos graves, aqueles com demandas com valores baixos pertencentes a classes com valores menores dos casos mais graves (que possuem demandas e pertencem às classes com valores mais altos).



## CONCLUSÃO

Aprendizado de máquina, inteligência artificial, modelos de raciocínio são temas extensos e que carecem ainda de bastante atenção por parte da comunidade acadêmica. São parte de um campo promissor de profundas transformações sociais, econômicas e culturais. Redes neurais, *big-data*, modelos de memória, entre outros, são cada dia mais utilizados em soluções comerciais e não somente no ramo científico.

Existem diversas considerações a serem levadas em conta caso se deseje aprofundar na matéria de Raciocínio Baseado em Casos para aprendizado de máquina, como por exemplo: o modelo de memória de casos, a granularidade da informação a ser apresentada, que casos devem ser representados, que estrutura pode ser mais facilmente adaptada a interface humano-computador, os casos de aquisição, representação, indexação, recuperação e os modelos de aprendizado concorrentes, entre outros, são temas que sugerimos serem explorados em trabalhos posteriores.

## REFERÊNCIAS BIBLIOGRÁFICAS

1. WANGENHEIM, C. G.; WANGENHEIM, A. e RATEKE, T. Raciocínio Baseado em Casos com Softwares Livres e Aplicativos Móveis. 2ª e. BOOKESS Editora, 2013.
2. WATSON, I D. Applying Case-Based Reasoning. Morgan Kauffman Publishers. San Francisco. 1997.
3. NAGY, H. M.; ALY, W. M. e HEGAZY, O. F. An Educational Data Mining System for Advising Higher Education Students. International Journal of Computer, Information Science and Engineering, Vol:7, N:10, 2013. World Academy of Science, Engineering and Technology, 2013
4. ABEL, M.; CASTILHO, J. M. V. Um Estudo Sobre Raciocínio Baseado em Casos. Porto Alegre, 1996. Trabalho para Pós-graduação em Ciência da Computação. Universidade Federal do Rio Grande do Sul. Porto Alegre, RS. 1996