

3

Arquitetura do Sistema

3.1.

Visão geral

O sistema desenvolvido permite a criação de aplicações que possibilitam efetuar consultas em um banco de dados relacional utilizando palavras-chave. O objetivo principal consiste em, a partir de uma fonte de dados, selecionar um determinado conjunto de objetos (tabelas e visões) que ficarão disponíveis para consultas. A criação da aplicação será feito por um usuário que tenha acesso ao banco de dados original, incluindo assim objetos e disponibilizando para outros usuários consultarem.

No momento da criação da aplicação, além de informar quais objetos ficarão acessíveis para consulta, o usuário poderá fornecer informações para melhorar o desempenho da aplicação. Uma senha permitirá o acesso, sempre que desejar, a uma área de administração da aplicação publicada com o objetivo de efetuar ajustes e administrar os objetos publicados.

Tabelas próprias do sistema irão armazenar as informações necessárias com o objetivo de minimizar o acesso às fontes de dados a todo o momento. Não haverá necessidade de acessar o banco de dados de origem a todo instante que uma consulta for submetida. A consulta à origem dos dados acontecerá apenas no momento que o usuário identificar qual consulta melhor atende à sua solicitação, diminuindo assim uma possível perda de desempenho no acesso a origem dos dados. Este conjunto de tabelas próprias do sistema pode ser reproduzida em qualquer sistema de gerenciamento de banco de dados.

3.2. Estrutura das tabelas do sistema

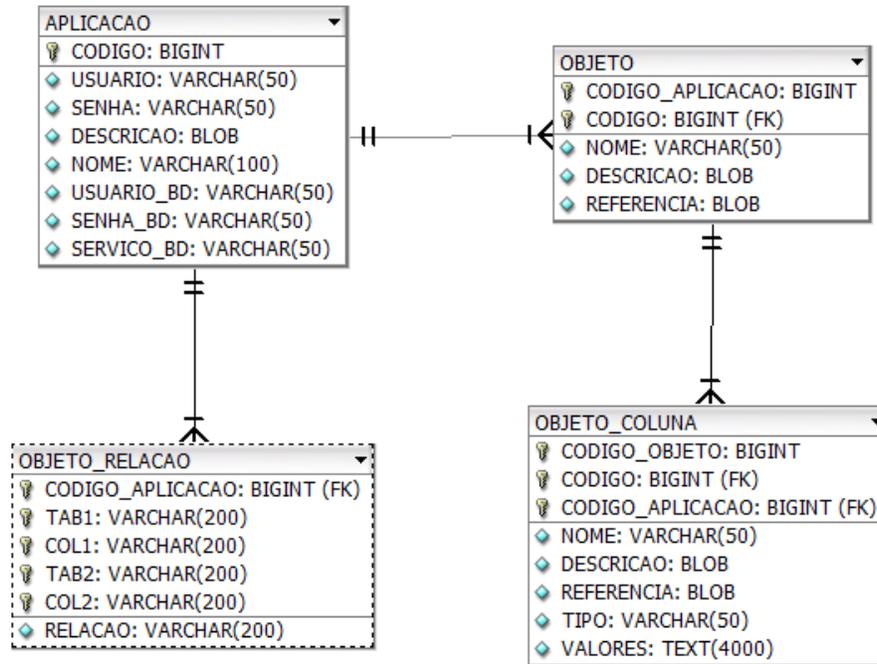


Figura 1: Estrutura do banco de dados do sistema

A Figura 1 apresenta a estrutura do banco de dados do sistema, os scripts de criação estão no anexo ao final deste documento. A seguir a descrição de cada tabela:

- **Aplicação:** Armazena o nome da aplicação, descrição, usuário, senha e conexão com o banco de dados de origem. Esta tabela armazena as informações iniciais quando o usuário cria uma aplicação (nome e descrição) bem como o usuário e senha de um usuário que pode alterar a aplicação (adicionando novas tabelas e efetuando configurações para melhorar as consultas).

- **Objeto:** Armazena os objetos (visões e tabelas) com seu nome e descrição das informações publicadas em uma determinada aplicação. Com esta tabela será possível encontrar os objetos que estão em uma determinada aplicação que será consultada sem a necessidade de consultar a origem dos dados.

- **Objeto Coluna:** Armazena as colunas dos objetos (visões e tabelas) inseridas na aplicação, especificamente na tabela Objeto. Além de armazenar o nome das colunas, coleta também a sua descrição, tipo e valores, no caso de campos textos. Com esta tabela é possível encontrar objetos e principalmente conteúdo textual, não sendo necessário efetuar consulta ao banco de dados original.

- **Objeto Relação:** Armazena as relações de todos os objetos inseridos na aplicação. Armazena as relações (chaves estrangeiras) entre tabelas incluindo suas colunas de relacionamento. Com esta tabela é possível encontrar o relacionamento entre as tabelas para construir as consultas sem a necessidade de consultar a origem dos dados.

Com este modelo é possível criar uma aplicação que apresenta um conjunto de tabelas publicadas para efetuar consultas por palavra-chave. Para cada aplicação, existe um conjunto de objetos (tabelas e visões) e metadados a eles associadas (colunas, relacionamentos, tipos, valores etc.), armazenados diretamente nas tabelas do sistema. Todas as relações das tabelas publicadas na aplicação ficam armazenadas para auxiliar na geração de consultas por palavra-chave, efetuando assim o relacionamento entre tabelas.

O sistema foi projetado para solicitar usuário e senha apenas da área de administração da aplicação. Porém, depois de publicada, qualquer usuário pode selecionar qual aplicação deseja consultar e submeter palavras-chave para recuperar informações.

3.3. Tipos de usuários

Existirão inicialmente dois níveis de usuários, um *usuário administrador* responsável pela criação (definição das tabelas que serão publicadas) e configuração da aplicação (adição de comentários e definições) e um *usuário comum*, que poderá realizar apenas consultas as aplicações já existentes.

O usuário administrador será responsável por acessar um módulo administrativo e publicar os objetos que farão parte da aplicação que será consultada pelo usuário. Ao efetuar esta publicação será necessário incluir principalmente comentários sobre cada coluna dos objetos inseridos. Estes metadados serão de extrema importância para efetuar a associação entre as palavras-chave e as colunas dos objetos do sistema.

A qualquer momento, em função de novas tabelas ou ajustes nos comentários existentes, o usuário administrador poderá retornar à área administrativa para efetuar novas configurações. Neste momento de configuração, o sistema efetuará requisições diretamente ao banco de dados de origem. Este processo de ajuste servirá para melhorar as respostas das consultas geradas caso estas não estejam sendo adequadas.

É importante que o usuário administrador da aplicação tenha conhecimento principalmente do conteúdo das tabelas que estão sendo inseridas. Porém, não necessita ter obrigatoriamente conhecimento técnico, mas apenas conceitual para disponibilizar um conteúdo para consulta. A inclusão de comentários contendo detalhes do objeto inserido é fundamental para que o sistema gere consultas em SQL automaticamente a partir de consultas por palavras-chave formuladas pelo usuário.

O usuário comum apenas consultará o sistema, não precisando conhecer nenhuma estrutura do banco de dados e nem ter conhecimento algum de linguagens de consulta. Ele precisará apenas informar palavras-chave, submetendo-as a uma aplicação selecionada. O usuário será solicitado a indicar quais associações geradas automaticamente pelo sistema estão corretas, auxiliando a geração das consultas.

3.4. Tecnologia utilizada

A conexão ao sistema é feita através de um navegador de Internet. Não será necessária nenhuma configuração adicional na máquina cliente que utilizará o sistema. As configurações serão efetuadas apenas no servidor Web.

Após acessar a página inicial, o usuário comum poderá submeter palavras-chave para consulta apenas escolhendo a aplicação que deseja consultar. O usuário administrador poderá acessar a página de administração para criar uma nova aplicação ou efetuar configurações em uma aplicação já existente (será necessário usuário e senha da aplicação).

A conexão do sistema instalado em um servidor Web com o banco de dados de origem e proprietária será feita utilizando OLEDB (uma biblioteca que permite o acesso a diversas fontes de dados).

No momento da criação de uma aplicação, o usuário administrador deverá fornecer o nome da base de dados que será consultada, assim como um usuário e senha que tenha permissão de acesso aos objetos que serão disponibilizados. O usuário comum não precisará saber nenhuma destas informações.

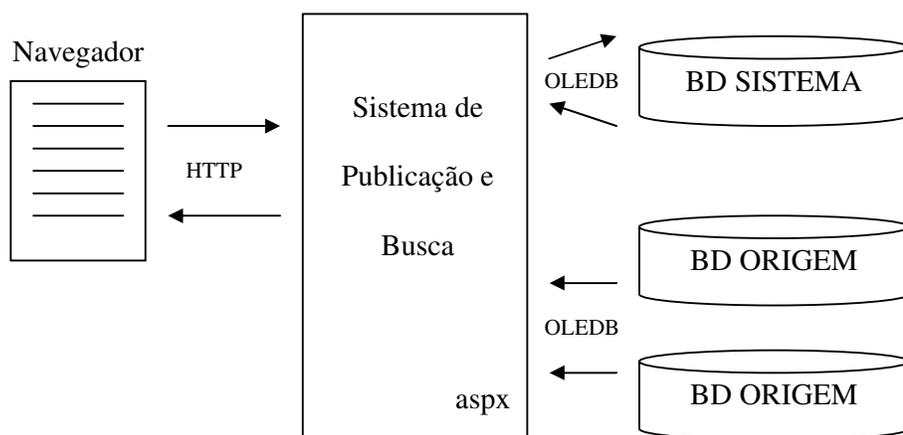


Figura 2: Funcionamento do sistema

O esquema da Figura 2 mostra que a conexão entre o navegador e o sistema é feito por HTTP para ambos os tipos de usuários. Para a criação de uma aplicação, o sistema se conecta ao banco de dados de origem e preenche as tabelas do banco de dados do sistema. Neste momento, as informações básicas, como nome das tabelas, e configurações adicionais que servirão de apoio para a geração automática das consultas são armazenadas dentro das tabelas próprias do sistema. Como já foi discutido, isto servirá para minimizar o acesso ao banco de dados de origem a todo o momento.

Nenhum dado é gravado ou alterado no banco de dados de origem. É preciso possuir apenas um usuário com permissão de leitura para os objetos que deseja disponibilizar.

O usuário que deseja efetuar uma busca irá consultar diretamente apenas o banco de dados do sistema para gerar as consultas, sem a necessidade de consultar a base de origem. A consulta à origem dos dados é feita apenas no momento de processar uma consulta selecionada pelo usuário.

No caso do sistema desenvolvido, somente será possível publicar objetos (tabelas ou visões) de uma fonte de dados que utilize banco de dados Oracle, porém a solução não apresenta nenhuma particularidade que impeça a utilização de qualquer outro sistema de gerência de bancos de dados relacional.

3.5. Etapas

Como já mencionado anteriormente, existem duas etapas bem distintas:

- **Publicação:** Etapa de pré-processamento, que consiste em fornecer os objetos, através de um módulo administrativo, que o sistema irá utilizar para processar consultas por palavra-chave. Estes objetos serão pré-processados para auxiliar a etapa de busca. Nesta etapa, as tabelas do sistema (aplicação, objeto, objeto coluna e objeto relação) são preenchidas com as informações coletadas do banco de dados.

- Busca: Etapa que consiste na entrada das palavras-chave, localização das informações relevantes no banco de dados publicado na fase anterior, criação das consultas em SQL e apresentação dos resultados.

A etapa de publicação é a etapa inicial quando se deseja disponibilizar objetos para consulta. Esta etapa deve ser feita por um usuário administrador que tenha acesso e permissão à origem dos dados. Após esta publicação, qualquer usuário poderá consultar os objetos previamente publicados sem a necessidade de conhecimento do banco de dados.

3.5.1. Etapa de Publicação

O sistema possui um conjunto de tabelas próprias responsáveis pelo armazenamento das informações necessárias no momento da busca. Estas informações são coletadas no momento que uma tabela é publicada no sistema.

Ao publicar uma tabela o sistema coleta os tipos, nomes e comentários de cada coluna permitindo ainda o usuário administrador fornecer novos comentários melhorando a capacidade de encontrar os resultados. O valor (conteúdo) das colunas do tipo texto também é coletado e armazenado nas tabelas do sistema. Isto é feito para que, no momento da geração das consultas, não ser necessário consultar a base de origem, melhorando o tempo de resposta.

As relações existentes entre a tabela publicada com outras tabelas também são levantadas e armazenadas no próprio sistema. Estas relações serão utilizadas para associar as palavras-chave no momento da busca. É muito importante fornecer uma boa descrição das colunas dos objetos, pois estas serão consultadas no momento da tentativa de associação entre as palavras-chave de entrada com as informações do banco de dados.

A qualquer momento é possível adicionar ou atualizar os objetos existentes na aplicação melhorando a geração automática das consultas. A Figura 3 mostra o fluxo que o sistema deverá executar para efetuar a criação de uma nova aplicação (publicação).

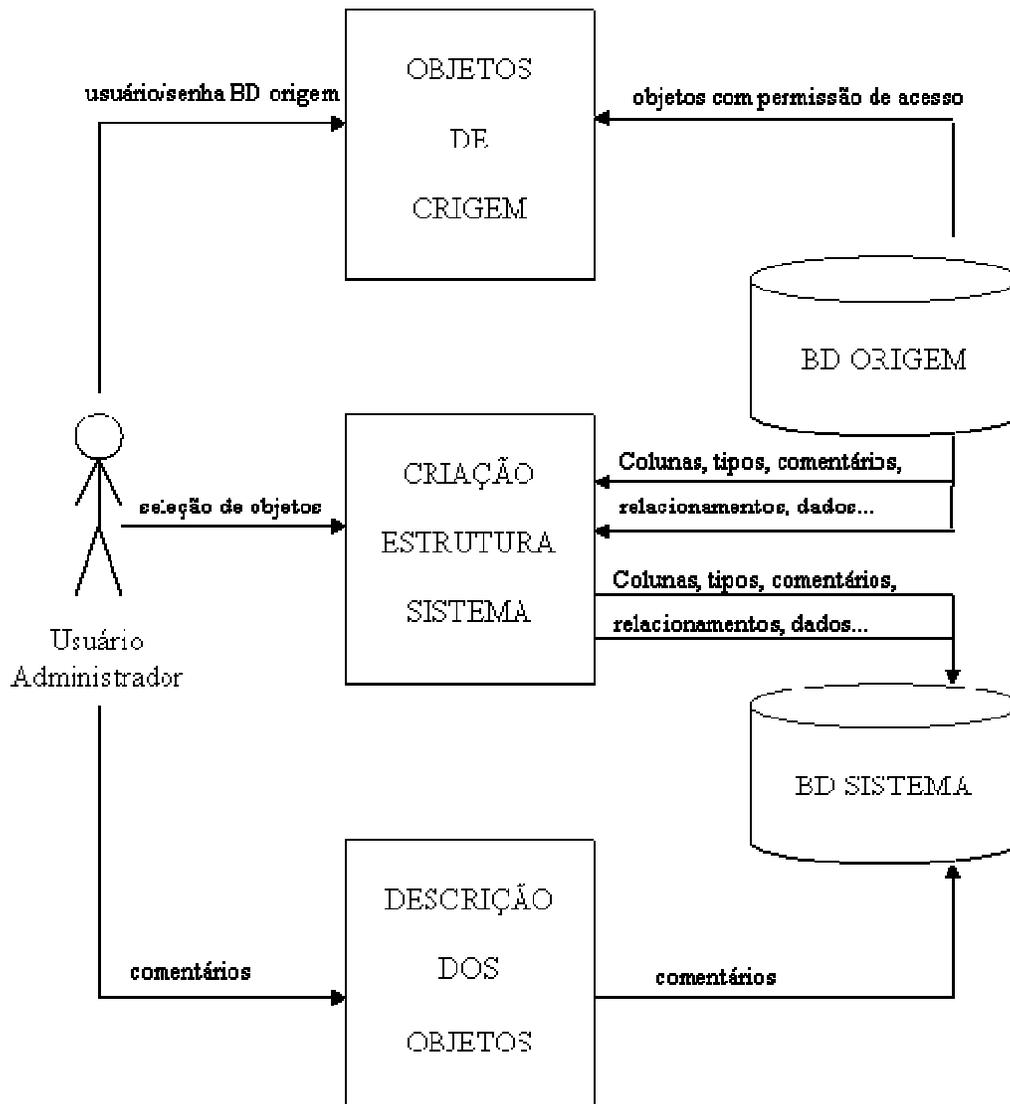


Figura 3: Etapa de publicação

Para exemplificar, digamos que desejamos criar uma aplicação para consulta a partir de duas tabelas (“Livro” e “Autor”) de um banco de dados relacional. O usuário administrador, ao entrar na parte de administração, poderá

selecionar estas duas tabelas que possui acesso (a partir do usuário e senha do banco de dados de origem).

Ao selecionar as tabelas “Livro” e “Autor” que deseja adicionar (seleção dos objetos) o sistema coleta do banco de dados de origem as colunas (nome do livro, código do livro, descrição do livro, nome do autor, código do autor e etc.), tipos (inteiros, texto e etc.), comentários e relacionamentos existentes das tabelas selecionadas. Estas informações são armazenadas no banco de dados do sistema para que seja utilizado no momento da geração da consulta.

O usuário administrador pode adicionar novos comentários como no código do livro escrever que seria o “ISBN”, assim se um usuário pesquisar utilizando a palavra-chave “ISBN” o sistema vai identificar que seria a coluna código do livro.

Os comentários servem também quando o nome das colunas é composto por nomes abreviados (siglas) como COD_L para o “código do livro” ou DESC_L para a “descrição do livro”.

3.5.2. Etapa de Busca

O usuário irá submeter ao sistema uma consulta não estruturada, isto é, conjunto de palavras-chave. O sistema transformará uma consulta não estruturada em uma consulta relacional cuja resposta está relacionada às palavras-chave fornecidas pelo usuário.

Após a entrada das palavras-chave, antes de efetuar a criação das consultas, o sistema irá consultar o conteúdo existente exclusivamente nas tabelas do sistema que foram preenchidas na etapa de Publicação para fornecer associações entre as palavras-chave de entrada e as colunas existentes no banco de dados.

Como é possível existir muitas associações, um resultado das possíveis associações será apresentado ao usuário para que ele possa selecionar quais associações parecem estar corretas. Com a seleção das melhores associações definidas pelo usuário o sistema irá efetuar a criação das consultas e apresentar possíveis resultados ao usuário. Como não é possível garantir a correta resposta a partir das palavras-chave fornecida como entrada pelo usuário, o sistema apresenta sugestões de resultados para que o usuário selecione o que realmente deseja consultar.

Os passos da etapa da busca são então:

- Entrada das palavras-chave efetuada pelo usuário.
- Recuperação das palavras-chave pelo sistema e busca nas tabelas do sistema, com o objetivo de identificar possíveis candidatos de associação entre uma palavra-chave e uma coluna de um objeto do banco de dados. É possível existir várias associações para uma mesma palavra-chave.
- Apresentação das possíveis associações efetuadas pelo sistema para o usuário selecionar as corretas.
- Seleção das possíveis associações relevantes aos usuários, servindo como um filtro inicial de resultados não relevantes.
- Recuperação das seleções efetuada pelo usuário e montagem de todos os conjuntos (enumeração) de pares palavra-chave e significado, formando assim conjuntos.
- Busca nas tabelas do sistema para identificar se existe relação entre os termos selecionados pelo usuário. Uma consulta será válida apenas quando os termos possuírem algum tipo de relação, caso contrário não será possível criar uma consulta com os mesmos.

- Geração das consultas apenas correspondentes aos conjuntos que apresentam relação entre todas as tabelas.
- Execução das consultas, gerando uma prévia do resultado para o usuário verificar qual gostaria de visualizar com detalhes.
- O usuário poderá selecionar quais são os melhores resultados, visualizando-os em detalhe.

Para exemplificar digamos que um usuário faça uma busca com as palavras-chave “autor ISBN 978-3-16-148410-0” (entrada de palavras-chave). O sistema identifica relações entre as palavras-chave de entrada com as colunas das tabelas que estão sendo pesquisadas e apresentar uma tabela com estas relações.

No caso “autor” será associado com a coluna “código do autor” e “nome do autor”, pois o sistema encontrou a palavra autor em cada uma das duas colunas, “ISBN” será associado com a coluna “código do livro”, pois esta coluna apresenta em sua descrição a palavra “ISBN” e por fim a “978-3-16-148410-0” será associado ao conteúdo do código do livro, pois será encontrado o seu valor nesta coluna.

Tendo esta tabela o usuário poderá marcar e desmarcar associações que não seja relevante, neste caso, por exemplo, o usuário poderia desmarcar a associação entre “autor” e “código do autor” caso esteja interessado apenas no nome do autor e não em seu código.

Na próxima etapa o sistema vai recuperar as seleções feitas pelo usuário e verificar se existe relação entre todos os termos escolhidos. Neste caso existe relação entre as duas tabelas envolvidas (“Livro” e “Autor”).

Uma consulta será gerada a partir do relacionamento existente entre as tabelas mostrando as colunas nome do autor e ISBN e filtrando o conteúdo da coluna “código do livro” que tenha “978-3-16-148410-0”.

A Figura 4 a seguir mostra o fluxo que o sistema deverá executar para efetuar a busca de uma aplicação existente.

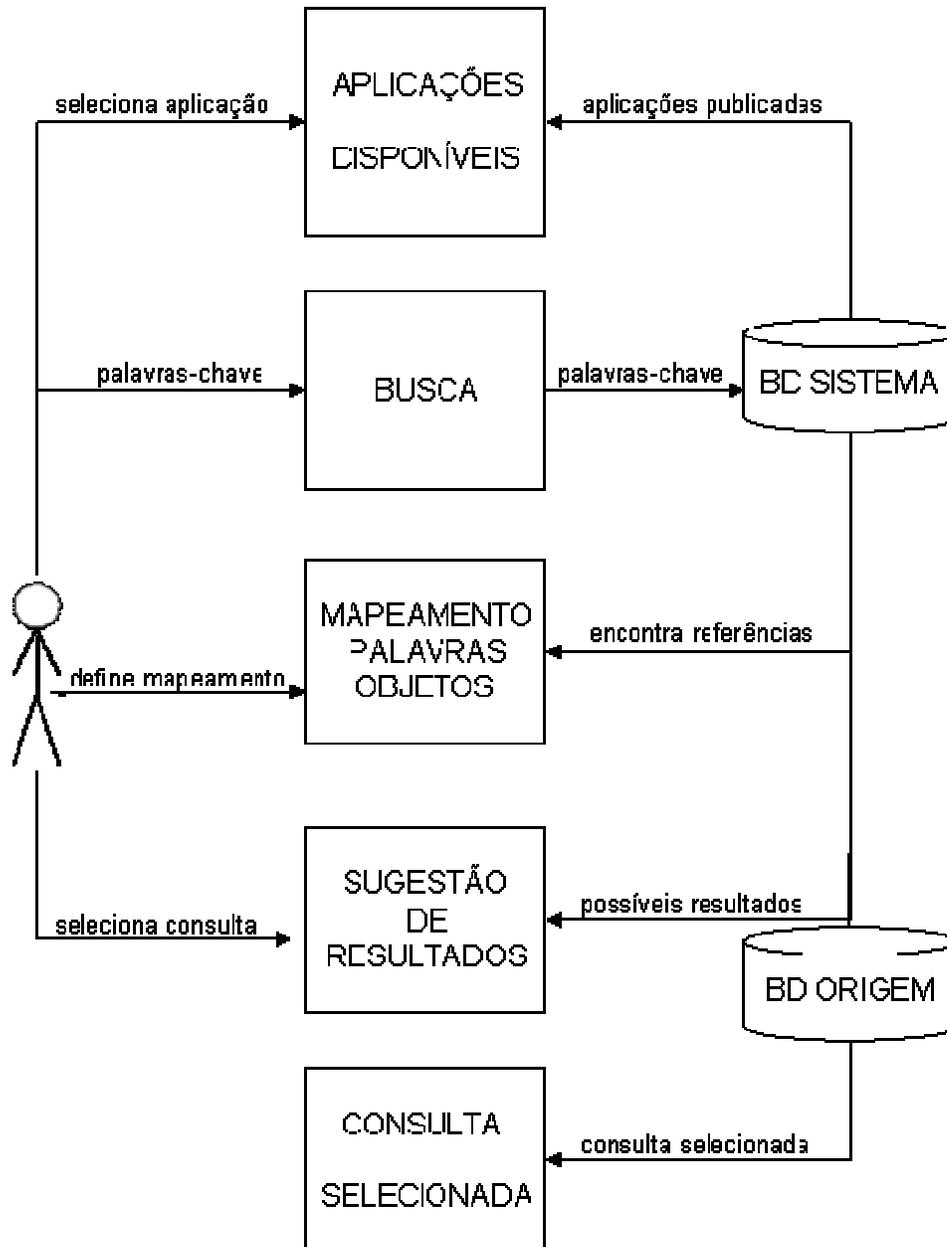


Figura 4: Etapa de busca