

2

Trabalhos Relacionados

2.1.

Introdução

Para dar suporte a este trabalho, alguns artigos foram consultados que servirão como fonte de informação. A seguir apresento alguns trabalhos relevantes.

2.2.

BANKS

Em “*Keyword Search in Databases*” [Hulgeri, Bhalotia, Nakhe & Chakrabarti, 2001], “*BANKS: Browsing and Keyword Searching in Relational Databases* [Aditya, Bhalotia, Chakrabarti, Hulgeri, Nakhe, Parag & Sudarshan, 2002] e “*Keyword Searching and Browsing in Databases using BANKS*” [Bhalotia, Hulgeri, Nakhe, Chakrabarti & Sudarshan, 2002], é apresentado o sistema chamado BANKS (*Browsing AND Keyword Searching*).

O sistema baseia-se na criação da estrutura de um grafo colocando as informações das linhas (tuplas) como nós e os relacionamentos entre as linhas como arestas de um grafo. Basicamente as relações chave primária com chave estrangeira são colocadas como arestas, porém outras arestas podem ser criadas para representar outros tipos de relações e dependências entre as linhas que não estejam contempladas nas chaves.

Pesos são atribuídos tanto as arestas quanto aos nós. Os pesos das arestas são representados pela proximidade entre eles, relevância semântica e tipo de ligação, já os pesos dos nós são atribuídos em função da sua relevância, quanto mais arestas apontar para um determinado nó, maior será a relevância. Estes pesos são utilizados para classificar as respostas encontradas.

Uma resposta a uma consulta é representada por um caminho onde cada nó representa uma relação com a palavra buscada e cada aresta uma ligação entre os elementos.

2.3. DBXplorer

Em “*DBXplorer A System for Keyword-Based Search over Relational Databases*” [Agrawal, Chaudhuri, & Das, 2002] é apresentado o sistema chamado DBXExplorer.

São adicionados recursos de busca por palavra-chave em bancos de dados comerciais, utilizando uma interface Web para exibir os resultados finais. O sistema apresenta uma etapa de pré-processamento que consiste basicamente da criação de uma tabela de símbolos e a etapa de busca aonde o mapeamento das palavras-chave com os símbolos são efetuados.

A tabela de símbolos pode ser criada de várias maneiras, sempre respeitando alguns critérios como espaço e tempo gasto para criação, desempenho da busca e facilidade de manutenção. O importante é existir um mapeamento fácil para encontrar qualquer coluna e tabela no banco de dados que esteja relacionado a palavra-chave.

Visando aperfeiçoar o desempenho e espaço uma estrutura *hash* é utilizada para encontrar rapidamente os valores e uma tabela de mapeamento para comprimir os dados. Um grafo é montado criando um conjunto de tabelas “relacionadas” que apresentam relevância com os termos buscados.

2.4. **DESANA**

Em “*DESANA: Uma Ferramenta para Extração de Dados da Web Considerando Contextos Fracos*” [Junior & Silva 2006] é apresentado o sistema DESANA (Data Extraction and Structural ANAlysis).

Este sistema implementa uma nova abordagem para geração semi-automática de extratores de dados de páginas da Web.

Com base em exemplos, padrões de extração são automaticamente inferidos para extração de valores de atributos mesmo quando o dado contextual nas páginas é pobre. Uma vez extraídos, os valores são agrupados para construir objetos complexos, de acordo como uma estrutura que é automaticamente inferida pela análise de suas posições nas páginas de origem.

A ferramenta dispõe de diversos recursos que facilitam a especificação de exemplos através de refinamentos sucessivos e testes interativos dos extratores sendo gerados.

2.5. **LABRADOR**

Em “*LABRADOR: Efficiently Publishing Relational Databases on the Web by using Keyword-Based Query Interfaces*” [Mesquita, Altigran, Moura, Calado & Laender, 2006] é apresentado o sistema LABRADOR.

O sistema efetua basicamente quatro passos. Primeiramente recebe as palavras-chave que representam a consulta não estruturada, depois gera consultas estruturadas candidatas através de um modelo de redes Bayesianas para estruturação de consultas, depois a partir da seleção do usuário o sistema gera uma consulta SQL equivalente, na etapa final a consulta é submetida ao banco de dados e ordenada utilizando novamente redes Bayesianas que estimam a probabilidade de uma tupla satisfazer a consulta do usuário.

O modelo Bayesiano realizado para estruturação das consultas gera um conjunto de pares ordenados que representam a associação dos termos procurados e os elementos existentes na base de dados. Note que cada elemento pode ocorrer mais de uma vez em uma consulta estruturada, porém cada termo buscado poderá aparecer apenas uma única vez. A ordenação, também usando redes Bayesianas, é calculada a partir das probabilidades em função da distribuição dos termos usados na consulta formulada pelo usuário ao longo da instância do banco de dados.

2.6. Comparativo

Cada trabalho apresenta suas particularidades para a geração das consultas a partir de palavras-chave. Estas particularidades variam desde a estrutura que será adotada para efetuar o mapeamento com a base de dados (*hash*, grafo, símbolos e etc.) quanto no critério adotado para classificação das consultas (distâncias, pesos, redes Bayesianas e etc.).

Esta dissertação irá apresentar outra nova maneira de estruturar os dados da base de dados de origem para que seja possível efetuar o mapeamento das palavras-chave.