

CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA

Disciplina: Administração de Banco de Dados

Full Text Searching

Professor: : José Antônio da Cunha

CEFET-RN

Full Text Searching

Uma necessidade cada vez mais comum de muitas aplicações, principalmente para aplicações web, é pesquisa por frases ou palavras não apenas nas colunas char ou varchar das tabelas de um banco de dados, mas também em arquivos como .txt, .pdf, .doc, .xls, entre outros.

O suporte à linguagem T-SQL fornecido pelo SQL Server permite que você realize pesquisa em dados armazenados em colunas do tipo char, nchar, varchar, nvarchar, text e ntext, mas os grandes problemas são:

O que fazer se você precisa pesquisar frases ou palavras em documentos?

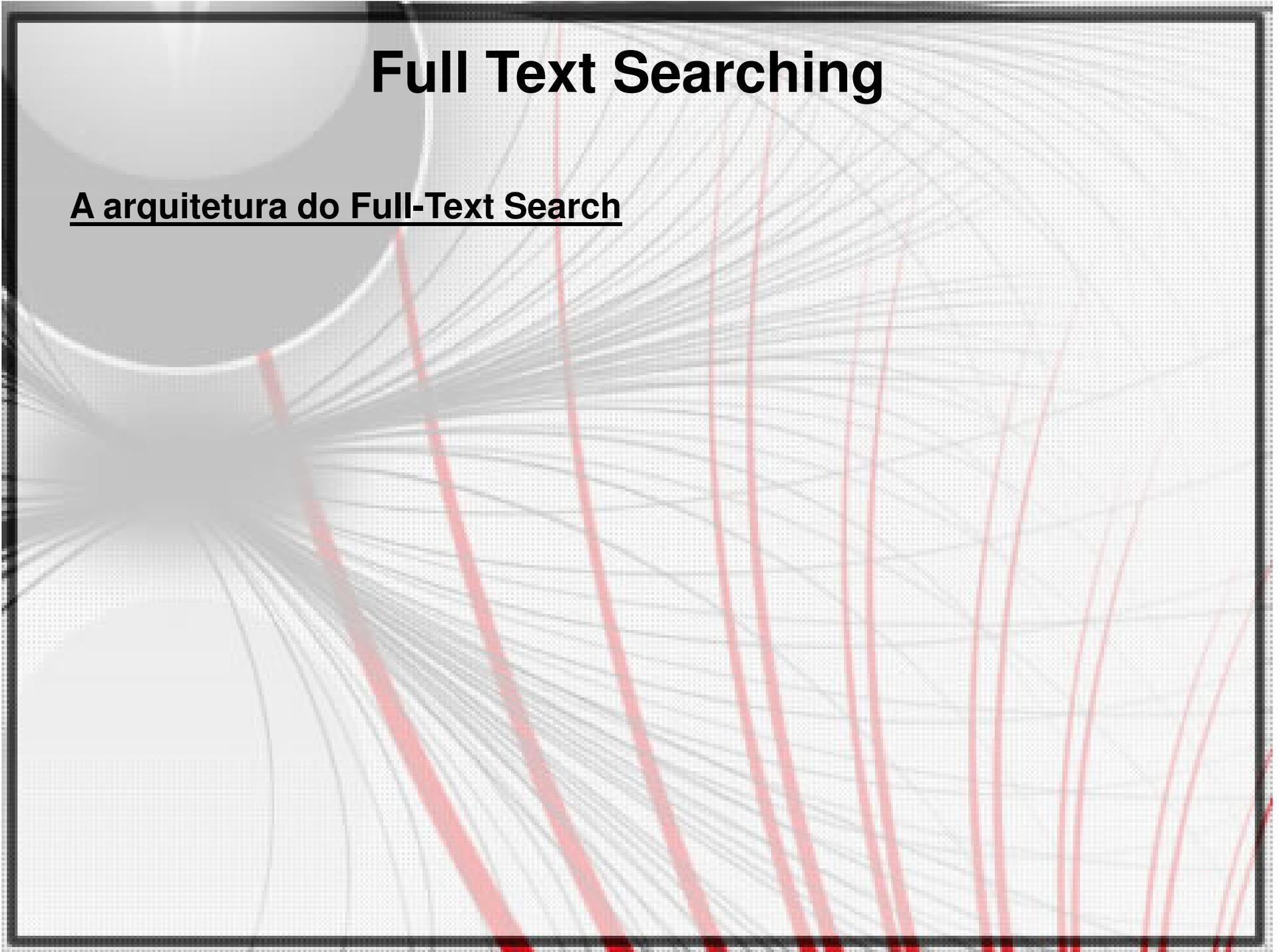
O que fazer se você precisa pesquisar não apenas uma palavra ou frase exata, como faz o LIKE, mas também por suas variantes verbais (exemplo: correr, correndo, correu) ou ainda estejam estas no singular ou no plural?

Full Text Searching

Presente no SQL Server desde sua versão 7.0, o Full-Text Search é a solução para estes problemas. Ele fornece suporte eficiente não apenas para busca de palavras ou frases em colunas baseadas em caracteres, mas também para a busca de palavras ou frases em arquivos .doc, .xls, .ppt, .pdf e vários outros

Full Text Searching

A arquitetura do Full-Text Search



Full Text Searching

O índice full-text (full-text index) é um tipo especial de índice que armazena informações sobre as palavras e suas respectivas localizações dentro de uma coluna. Estas informações são posteriormente utilizadas pelo full-text search para otimizar as consultas full-text que buscam por uma palavra específica ou uma combinação delas. Estes índices são armazenados em estruturas conhecidas como catálogo (full-text catalogs) e juntos são mantidos como uma coleção de pastas e arquivos no disco local do seu servidor.

Os catálogos podem armazenar índices full-text para uma ou mais tabelas desde que estas tabelas pertençam a um único banco de dados, ou seja, um catálogo não pode pertencer a múltiplos bancos de dados.

Full Text Searching

Você pode armazenar até 2GB de texto em um único campo no SQL Server. Este conteúdo será então indexado e utilizado pelo mecanismo de busca do Microsoft Search Service. O principal requerimento para implementação de um Full-Text é a criação de um índice em uma coluna simples em todas as tabelas selecionadas para Full-Text Search.

Full Text Searching

Quando o SQL Server recebe uma query, que exige full-text search, ele envia o critério de seleção para o Microsoft Search Service, que faz a busca e retorna um valor chave e um ranking de valores para cada linha correspondente. O MSSQLServer service utiliza essa informação para montar o result set da query.

Para evitar que haja problemas durante o processo de busca, algumas palavras existentes no meio de uma frase serão eliminadas, pois não tem nenhum significado, como “and”, “is”, etc.

Vale ressaltar que a lista de palavras pode ser customizada.

Full Text Searching

Implementando Full-Text Search

Antes de executarmos alguma query é necessário verificar se o serviço Microsoft Search está rodando. Devemos criar índices de full-text nas tabelas que serão pesquisadas.

Podemos iniciar ou parar o serviço Microsoft Search das seguintes maneiras:

- Usando o menu de contexto do objeto Full-Text Search no SQL Server Enterprise Manager.
- Usando o Server Service Manager e selecionando Microsoft Search.
- Executando `net start mssearch` (ou `net stop mssearch`) do command prompt.

Full Text Searching

Criando Full-Text Index

Devemos considerar dois pontos importantes quando trabalhamos com Full-Text: a capacidade de realizarmos queries diretamente com campos do tipo caractere e a necessidade de atualização dos índices.

Quando trabalhamos com full-text indexes, devemos considerar os seguintes pontos:

Full Text Searching

- ❑ Índices full-text são armazenados no file system, mas gerenciados pelo database;
- ❑ Existe apenas um único índice por tabela;
- ❑ A adição de dados nos índices será feita por um agendamento ou requisição de tarefa;
- ❑ Os índices existentes em um database são agrupados em um único full-text catalog;
- ❑ É interessante criarmos catálogos diferentes para tabelas muito grandes, isso reduz o tempo de repopular os índices existentes em tabelas menores, pois estarão em tabelas separadas.

Full Text Searching

Populando Índices de Full-Text

Podemos popular os índices de duas maneiras diferentes:

Full Population – Esse método atualizará todos os índices de um catálogo, independentemente de ter havido, ou não, mudanças no conteúdo dos registros desde a última tarefa de população. Usamos normalmente este método ao popular um catálogo pela primeira vez.

Full Text Searching

Populando Índices de Full-Text

Incremental Population – Esse método atualizará os índices de acordo com as alterações que foram realizadas nos registros desde a última tarefa de população. Esta população de dados se tornará uma FULL POPULATION, automaticamente, se ocorrer uma das seguintes situações:

- Uma tabela sem uma coluna timestamp for colocada para full-text indexing
- Novas colunas forem colocadas para processamento full-text desde a última população de dados
- A estrutura da tabela tenha sido alterada desde a última população

Full Text Searching

Atualizando Índices de Full-Text

Diferentemente dos índices de um banco de dados relacional, os índices de full-text, não são atualizados instantaneamente, quando ocorre uma atualização de dados ou quando linhas são inseridas nas tabelas registradas para full-text, nem mesmo quando essas linhas são deletadas.

Full Text Searching

Atualizando Índices de Full-Text

O processo de população de dados deve ser iniciado manualmente ou agendado para ocorrer em intervalos pré-estabelecidos. Esses índices são populados de forma assíncrona pelas seguintes razões:

- É necessário muito mais tempo para atualizar um full-text index, do que normalmente é necessário para um índice relacional
- As pesquisas de full-text são normalmente menos precisas que as pesquisas padrão. Então, a necessidade de um sincronismo dinâmico não é prioritária

Full Text Searching

Adicionando Full-Text Search em uma Tabela

1. Abra o SQL Server Enterprise Manager.
2. Expandir o console tree, Databases, northwind e clique em tables.
3. No painel de detalhes, selecione a tabela employees e clique com o botão direito, no menu, selecione Full-Text Index Table, e clique Define Full-Text Indexing On A Table.

Full Text Searching

4. Use as seguintes informações para configuração:

Opção	Valor
Select a unique index	PK_Employee
Added columns	Notes
Create a new catalog?	Checked
New catalog _ Name	Northwind_catalog
Select or Create Population Schedules (Optional)	No

Full Text Searching

5. Você receberá a informação da criação do índice, mas será informado que o mesmo deve ser populado.
6. Abra SQL Server Query Analyzer, e de um logon para (local) server com Microsoft Windows NT authentication.
7. Execute a seguinte stored procedure para confirmar que o índice foi criado:

```
USE Northwind  
EXEC sp_help_fulltext_tables
```

Full Text Searching

Criando Índices Full-Text na prática:

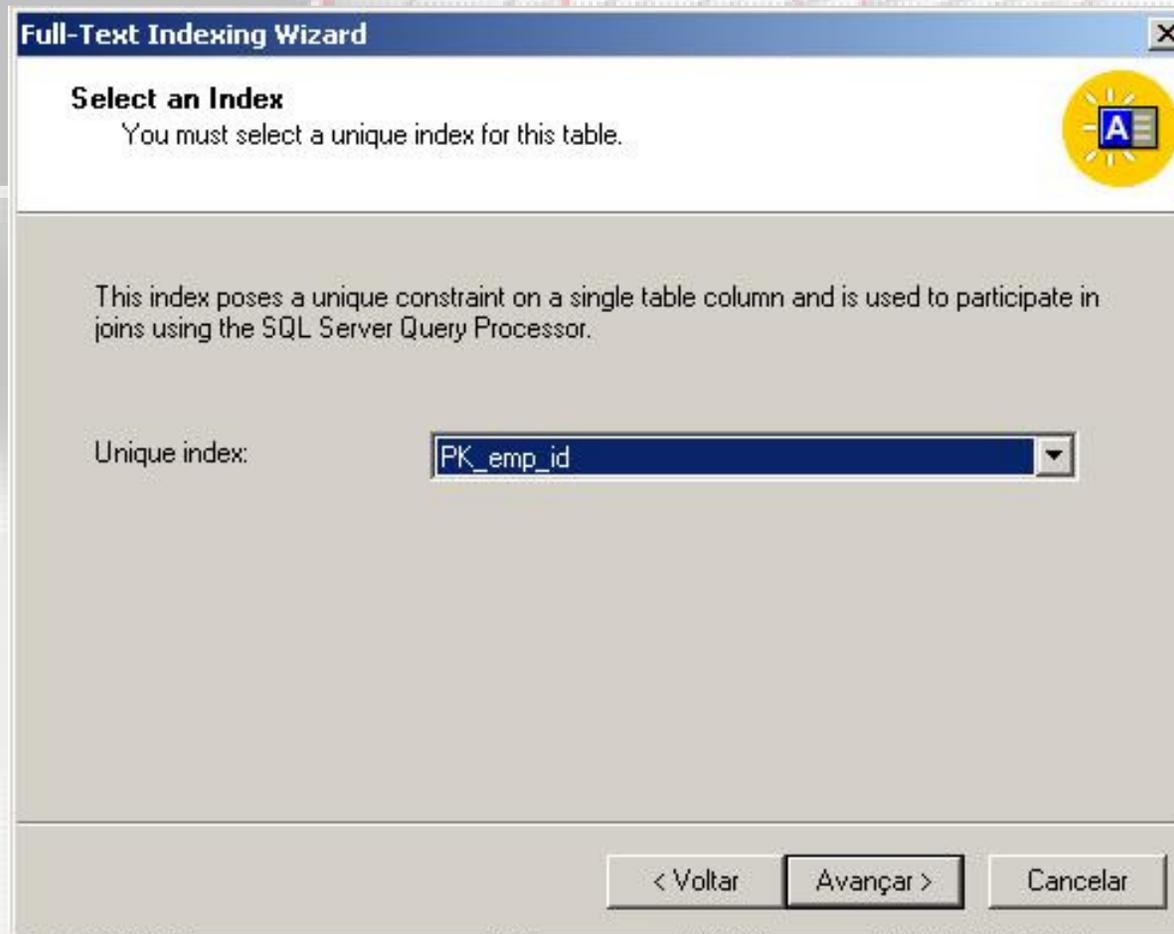
No Enterprise Manager, selecione o Banco e escolha a tabela para inserir um Índice, conforme a figura, depois clique em avançar.

The screenshot displays the Microsoft SQL Server Enterprise Manager interface. On the left, a tree view shows the 'pubs' database expanded to the 'Tables' folder. The main pane shows a list of tables in the 'pubs' database, including 'discounts', 'employee', 'jobs', 'pub_info', 'publishing', 'roysched', 'sales', 'stores', 'syscolumns', 'syscomments', 'sysdependencies', 'sysfilegroups', 'sysfiles', 'sysfiles2', 'sysfiles3', 'sysfiles4', 'sysfiles5', 'sysfiles6', 'sysfiles7', 'sysfiles8', 'sysfiles9', 'sysfiles10', 'sysfiles11', 'sysfiles12', 'sysfiles13', 'sysfiles14', 'sysfiles15', 'sysfiles16', 'sysfiles17', 'sysfiles18', 'sysfiles19', 'sysfiles20', 'sysfiles21', 'sysfiles22', 'sysfiles23', 'sysfiles24', 'sysfiles25', 'sysfiles26', 'sysfiles27', 'sysfiles28', 'sysfiles29', 'sysfiles30', 'sysfiles31', 'sysfiles32', 'sysfiles33', 'sysfiles34', 'sysfiles35', 'sysfiles36', 'sysfiles37', 'sysfiles38', 'sysfiles39', 'sysfiles40', 'sysfiles41', 'sysfiles42', 'sysfiles43', 'sysfiles44', 'sysfiles45', 'sysfiles46', 'sysfiles47', 'sysfiles48', 'sysfiles49', 'sysfiles50', 'sysfiles51', 'sysfiles52', 'sysfiles53', 'sysfiles54', 'sysfiles55', 'sysfiles56', 'sysfiles57', 'sysfiles58', 'sysfiles59', 'sysfiles60', 'sysfiles61', 'sysfiles62', 'sysfiles63', 'sysfiles64', 'sysfiles65', 'sysfiles66', 'sysfiles67', 'sysfiles68', 'sysfiles69', 'sysfiles70', 'sysfiles71', 'sysfiles72', 'sysfiles73', 'sysfiles74', 'sysfiles75', 'sysfiles76', 'sysfiles77', 'sysfiles78', 'sysfiles79', 'sysfiles80', 'sysfiles81', 'sysfiles82', 'sysfiles83', 'sysfiles84', 'sysfiles85', 'sysfiles86', 'sysfiles87', 'sysfiles88', 'sysfiles89', 'sysfiles90', 'sysfiles91', 'sysfiles92', 'sysfiles93', 'sysfiles94', 'sysfiles95', 'sysfiles96', 'sysfiles97', 'sysfiles98', 'sysfiles99', 'sysfiles100'. A context menu is open over the 'employee' table, with the 'Full-Text Index Table' option selected. The 'Define Full-Text Indexing on a Table...' option is highlighted in the sub-menu.

Table Name	Schema	User	Created
discounts	pubs	User	6/8/2000 01:33:52
employee	pubs	User	6/8/2000 01:33:53
jobs	pubs	User	6/8/2000 01:33:52
pub_info	pubs	User	6/8/2000 01:33:53
publishing	pubs	User	6/8/2000 01:33:52
roysched	pubs	User	6/8/2000 01:33:52
sales	pubs	User	6/8/2000 01:33:52
stores	pubs	User	6/8/2000 01:33:52
syscolumns	pubs	User	6/8/2000 01:33:52
syscomments	pubs	User	6/8/2000 01:33:52
sysdependencies	pubs	User	6/8/2000 01:33:52
sysfilegroups	pubs	User	6/8/2000 01:33:52
sysfiles	pubs	User	6/8/2000 01:33:52
sysfiles2	pubs	User	6/8/2000 01:33:52
sysfiles3	pubs	User	6/8/2000 01:33:52
sysfiles4	pubs	User	6/8/2000 01:33:52
sysfiles5	pubs	User	6/8/2000 01:33:52
sysfiles6	pubs	User	6/8/2000 01:33:52
sysfiles7	pubs	User	6/8/2000 01:33:52
sysfiles8	pubs	User	6/8/2000 01:33:52
sysfiles9	pubs	User	6/8/2000 01:33:52
sysfiles10	pubs	User	6/8/2000 01:33:52
sysfiles11	pubs	User	6/8/2000 01:33:52
sysfiles12	pubs	User	6/8/2000 01:33:52
sysfiles13	pubs	User	6/8/2000 01:33:52
sysfiles14	pubs	User	6/8/2000 01:33:52
sysfiles15	pubs	User	6/8/2000 01:33:52
sysfiles16	pubs	User	6/8/2000 01:33:52
sysfiles17	pubs	User	6/8/2000 01:33:52
sysfiles18	pubs	User	6/8/2000 01:33:52
sysfiles19	pubs	User	6/8/2000 01:33:52
sysfiles20	pubs	User	6/8/2000 01:33:52
sysfiles21	pubs	User	6/8/2000 01:33:52
sysfiles22	pubs	User	6/8/2000 01:33:52
sysfiles23	pubs	User	6/8/2000 01:33:52
sysfiles24	pubs	User	6/8/2000 01:33:52
sysfiles25	pubs	User	6/8/2000 01:33:52
sysfiles26	pubs	User	6/8/2000 01:33:52
sysfiles27	pubs	User	6/8/2000 01:33:52
sysfiles28	pubs	User	6/8/2000 01:33:52
sysfiles29	pubs	User	6/8/2000 01:33:52
sysfiles30	pubs	User	6/8/2000 01:33:52
sysfiles31	pubs	User	6/8/2000 01:33:52
sysfiles32	pubs	User	6/8/2000 01:33:52
sysfiles33	pubs	User	6/8/2000 01:33:52
sysfiles34	pubs	User	6/8/2000 01:33:52
sysfiles35	pubs	User	6/8/2000 01:33:52
sysfiles36	pubs	User	6/8/2000 01:33:52
sysfiles37	pubs	User	6/8/2000 01:33:52
sysfiles38	pubs	User	6/8/2000 01:33:52
sysfiles39	pubs	User	6/8/2000 01:33:52
sysfiles40	pubs	User	6/8/2000 01:33:52
sysfiles41	pubs	User	6/8/2000 01:33:52
sysfiles42	pubs	User	6/8/2000 01:33:52
sysfiles43	pubs	User	6/8/2000 01:33:52
sysfiles44	pubs	User	6/8/2000 01:33:52
sysfiles45	pubs	User	6/8/2000 01:33:52
sysfiles46	pubs	User	6/8/2000 01:33:52
sysfiles47	pubs	User	6/8/2000 01:33:52
sysfiles48	pubs	User	6/8/2000 01:33:52
sysfiles49	pubs	User	6/8/2000 01:33:52
sysfiles50	pubs	User	6/8/2000 01:33:52
sysfiles51	pubs	User	6/8/2000 01:33:52
sysfiles52	pubs	User	6/8/2000 01:33:52
sysfiles53	pubs	User	6/8/2000 01:33:52
sysfiles54	pubs	User	6/8/2000 01:33:52
sysfiles55	pubs	User	6/8/2000 01:33:52
sysfiles56	pubs	User	6/8/2000 01:33:52
sysfiles57	pubs	User	6/8/2000 01:33:52
sysfiles58	pubs	User	6/8/2000 01:33:52
sysfiles59	pubs	User	6/8/2000 01:33:52
sysfiles60	pubs	User	6/8/2000 01:33:52
sysfiles61	pubs	User	6/8/2000 01:33:52
sysfiles62	pubs	User	6/8/2000 01:33:52
sysfiles63	pubs	User	6/8/2000 01:33:52
sysfiles64	pubs	User	6/8/2000 01:33:52
sysfiles65	pubs	User	6/8/2000 01:33:52
sysfiles66	pubs	User	6/8/2000 01:33:52
sysfiles67	pubs	User	6/8/2000 01:33:52
sysfiles68	pubs	User	6/8/2000 01:33:52
sysfiles69	pubs	User	6/8/2000 01:33:52
sysfiles70	pubs	User	6/8/2000 01:33:52
sysfiles71	pubs	User	6/8/2000 01:33:52
sysfiles72	pubs	User	6/8/2000 01:33:52
sysfiles73	pubs	User	6/8/2000 01:33:52
sysfiles74	pubs	User	6/8/2000 01:33:52
sysfiles75	pubs	User	6/8/2000 01:33:52
sysfiles76	pubs	User	6/8/2000 01:33:52
sysfiles77	pubs	User	6/8/2000 01:33:52
sysfiles78	pubs	User	6/8/2000 01:33:52
sysfiles79	pubs	User	6/8/2000 01:33:52
sysfiles80	pubs	User	6/8/2000 01:33:52
sysfiles81	pubs	User	6/8/2000 01:33:52
sysfiles82	pubs	User	6/8/2000 01:33:52
sysfiles83	pubs	User	6/8/2000 01:33:52
sysfiles84	pubs	User	6/8/2000 01:33:52
sysfiles85	pubs	User	6/8/2000 01:33:52
sysfiles86	pubs	User	6/8/2000 01:33:52
sysfiles87	pubs	User	6/8/2000 01:33:52
sysfiles88	pubs	User	6/8/2000 01:33:52
sysfiles89	pubs	User	6/8/2000 01:33:52
sysfiles90	pubs	User	6/8/2000 01:33:52
sysfiles91	pubs	User	6/8/2000 01:33:52
sysfiles92	pubs	User	6/8/2000 01:33:52
sysfiles93	pubs	User	6/8/2000 01:33:52
sysfiles94	pubs	User	6/8/2000 01:33:52
sysfiles95	pubs	User	6/8/2000 01:33:52
sysfiles96	pubs	User	6/8/2000 01:33:52
sysfiles97	pubs	User	6/8/2000 01:33:52
sysfiles98	pubs	User	6/8/2000 01:33:52
sysfiles99	pubs	User	6/8/2000 01:33:52
sysfiles100	pubs	User	6/8/2000 01:33:52

Full Text Searching

Clique em avançar na próxima tela.



The image shows a screenshot of a Windows-style dialog box titled "Full-Text Indexing Wizard". The dialog has a blue title bar with a close button (X) in the top right corner. The main content area is light gray and contains the following text:

Select an Index
You must select a unique index for this table.

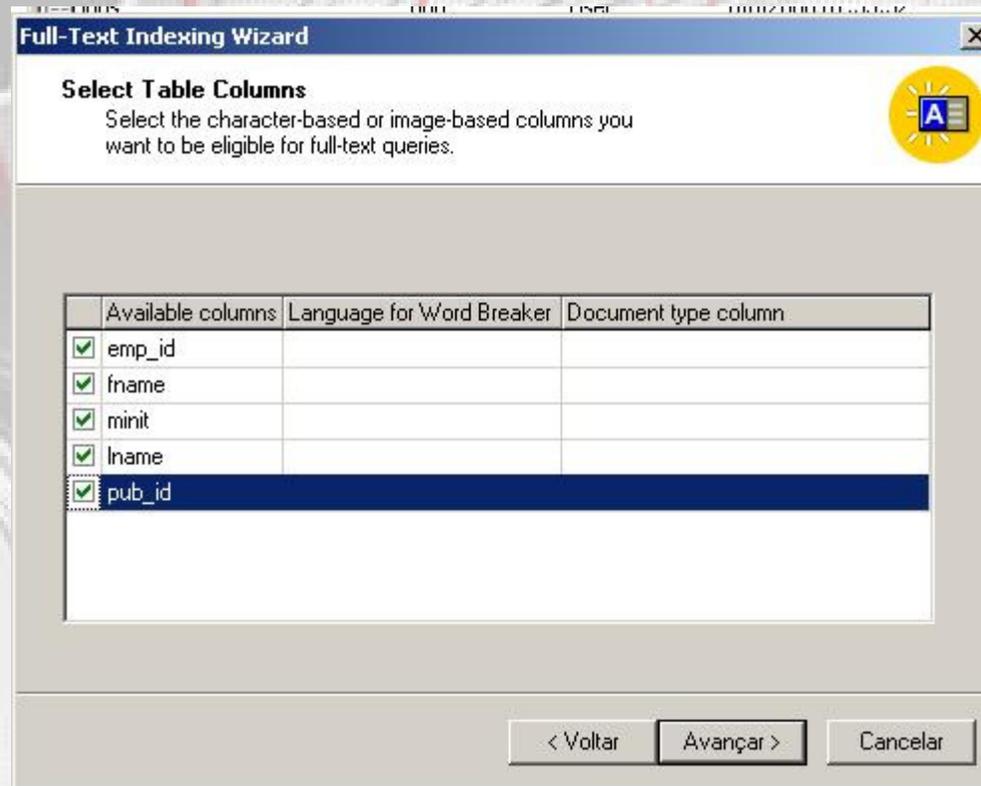
This index poses a unique constraint on a single table column and is used to participate in joins using the SQL Server Query Processor.

Unique index:

At the bottom of the dialog, there are three buttons: "< Voltar", "Avançar >", and "Cancelar". A yellow circular icon with a blue 'A' and a sunburst is located in the top right corner of the main content area.

Full Text Searching

Selecione as colunas que você quer que façam parte da Pesquisa full-text.



Full Text Searching

Escolha o nome do catalogo:

Full-Text Indexing Wizard

Select a Catalog
You must select an existing full-text catalog or create one for this database.

If this is a large table, or if the chosen full-text catalog has indexed other large tables, consider assigning this table to its own full-text catalog.

Select full-text catalog: No catalogs available

Create a new catalog

New catalog

Name: Northwind_catalog

Location: C:\Arquivos de programas\Microsoft SQL Ser

< Voltar Avançar > Cancelar

Full Text Searching

Full-Text Indexing Wizard [X]

Select or Create Population Schedules (Optional)
You can manage the population of this full-text catalog by creating a new catalog or table population schedule or modifying an existing one.



Name	Type	Enabled	Description

New Iable Schedule... New Catalog Schedule... Edit Delete

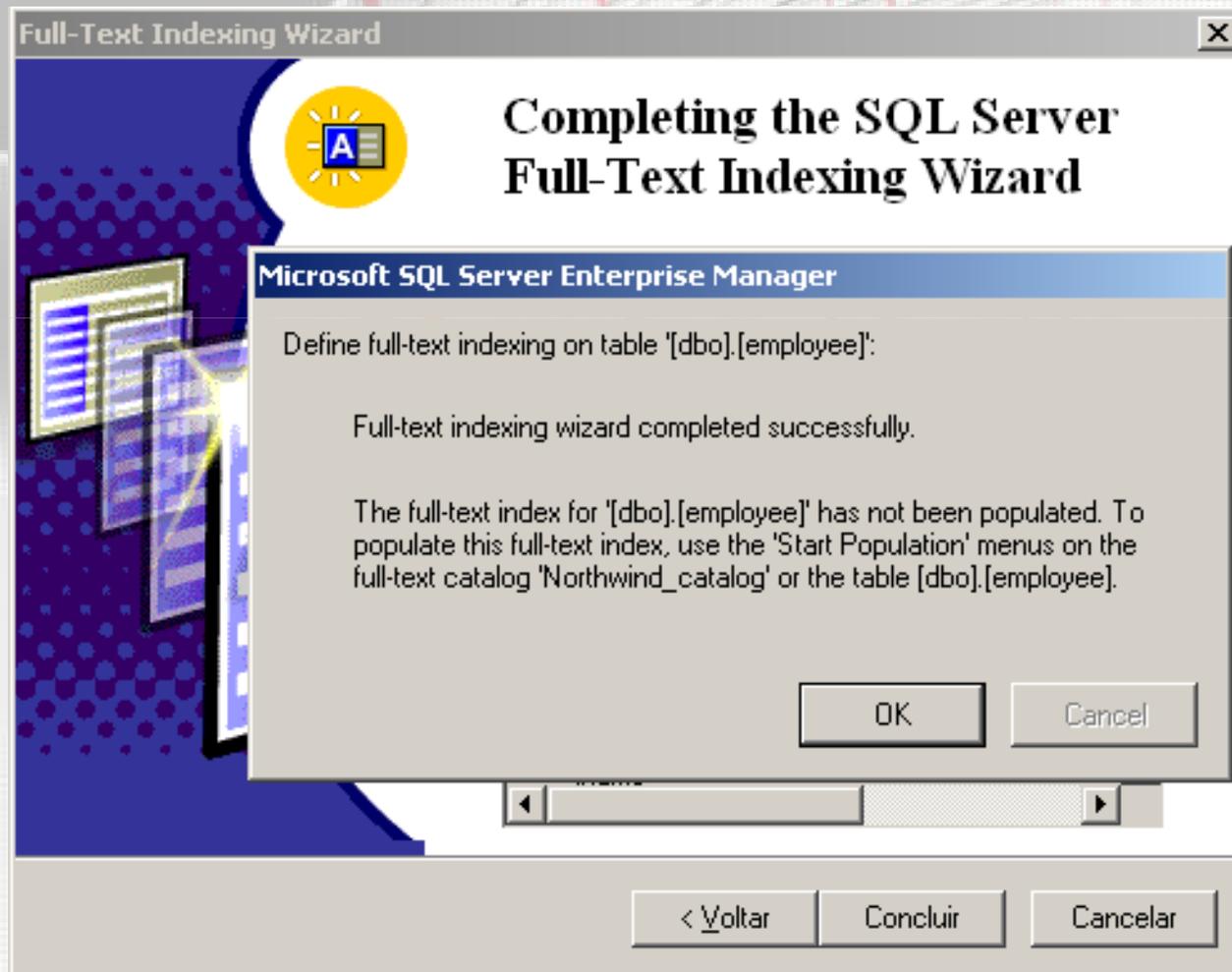
< Voltar Avançar > Cancelar

Full Text Searching

Clique em concluir para finalizar.



Full Text Searching



Full Text Searching

Populando o catálogo para full-text searching.

1. Vá para o SQL Server Enterprise Manager.
2. No console tree, clique em Full-Text Catalogs no Pubs database.
3. No painel de detalhes, clique com o botão direito em Northwind_catalog e clique em Properties. Veja as informações sobre o catálogo. Perceba que o status de população é o idle corrente e o item count é 0. Feche a janela de propriedades.
4. Clique com o botão direito em Northwind_catalog, selecione Start Population e clique em Full Population para popular o full-text index.
5. Clique com o botão direito em Northwind_catalog e selecione Properties. Veja as informações sobre o catálogo. Se você abrir a janela de propriedades rapidamente, você poderá ainda ver que o status de população permanece em “PROGRESS”.

Full Text Searching

Construindo Full-Text Queries

Com o uso de full-text query, você poderá executar funções de busca avançada em campos de texto nas tabelas marcadas para full-text. Diferentemente do operador LIKE, que é utilizado para busca de caracteres, o full-text opera através de uma combinação de palavras e frases, reportando a compatibilidade dos resultados obtidos pela busca.

Usando Transact-SQL Predicates e Funções

Você pode utilizar os seguintes Transact-SQL predicates e funções row-set para escrever full-text queries:

- Usando CONTAINS e FREETEXT predicates nas condições de busca (incluindo a cláusula WHERE) de um SELECT.
- Usando as funções CONTAINSTABLE e FREETEXTTABLE na cláusula FROM de um SELECT.

Full Text Searching

Sintaxe:

```
CONTAINS({column | *}, '<contains_search_condition>')
```

```
FREETEXT({column | * }, 'freetext_string')
```

```
CONTAINSTABLE(table, {column | *}, '<contains_search_condition>')
```

```
FREETEXTTABLE (table, {column | *}, 'freetext_string')
```

Exemplos:

A query a seguir retorna o `plant_id`, `common_name` e `price` para todas as linhas que tenham a frase "English Thyme" em qualquer uma das colunas de full-text.

```
SELECT plant_id, common_name, price FROM plants  
WHERE CONTAINS( *, ' "English Thyme" ' )
```

Full Text Searching

A query a seguir retorna linhas nas quais a descrição indexada possui o texto "Jean LeDuc has always loved ice hockey" ou "Jean Leduc on Ice—Hockey at Its Best."

```
SELECT article_id FROM hockey_articles  
WHERE CONTAINS (description, ' "Jean LeDuc" AND "ice hockey" ' )
```

O exemplo a seguir, utiliza a FREETEXT predicate em uma coluna chamada description.

```
SELECT * FROM news_table WHERE FREETEXT  
(description, ' "The Fulton County Grand Jury said  
Friday an investigation of Atlanta's recent primary  
election produced no evidence that any irregularities took place." ' )
```

Full Text Searching

Vamos executar então algumas queries que utilizem o Full-Text Search.

1. Abra o SQL Server Query Analyzer.
2. Execute as seguintes queries:

```
USE Northwind
SELECT lastname, title, hiredate, notes
FROM employees
WHERE CONTAINS (notes, "sales management")
```

```
USE Northwind
SELECT lastname, title, hiredate, notes
FROM employees
WHERE CONTAINS (notes, "sales" AND "management")
```

```
USE Northwind
SELECT lastname, title, hiredate, notes
FROM employees
WHERE CONTAINS (notes, "sales" NEAR "management")
```

Full Text Searching

F I M