

Linguagens Formais e Autômatos

Laboratório de Engenharia de Gramáticas



Objetivo: Conhecer um ambiente de desenvolvimento de gramáticas e parsing

Sistema: LKB <http://wiki.delph-in.net/moin/LkbTop>

Atividade: Estender uma gramática e fazer um relatório de no máximo 1 página descrevendo os resultados obtidos. Submeta o relatório no moodle até o dia 27/04.

Roteiro:

I - Instalando e Iniciando o LKB

- 1) Abra uma janela do Internet Explorer ou Firefox
 - a. Abra a locação: <http://wiki.delph-in.net/moin/LkbTop>
 - b. Percorre a página até a seção **Background**
 - c. Selecione: **[LkbInstallation: Obtaining, Installing and Running the LKB](#)**;
 - d. Percorre a página até a seção **Quick Installation on Windows**
 - e. Clique no link indicado no passo 1 (Download the current Windows version of the  **LKB** (2004-09-17)) para fazer o download do LKB
 - f. Salve o arquivo lkb_windows.zip num diretório e o descompacte
 - g. Isto irá criar o diretório **windows** que contém entre outros arquivos **lkb.exe**.
 - h. Faça o download também das gramáticas seguindo o passo 3 (Download the  **grammar files**)
 - i. Salve o arquivo lkb_data.tgz e usando 7-zip o descompacte no diretório **windows** (criado na descompactação de lkb_windows.zip no passo f acima)
 - j. Isto criará o diretório **src** contendo o diretório **data** contendo uma variedade de gramáticas
 - k. Crie um novo diretório em **C:** chamado **tmp**.
- 2) Vá para o diretório **windows**
- 3) Clique em lkb.exe
 - a. A janela TOP do LKB será ativada
 - b. Do menu principal selecione **Load**
 - i. Selecione **Complete Grammar**
 - ii. Do diretório
`windows/src/data/esslli2k/grammars/grammar1`
selecione o arquivo **script**
 - c. Uma janela com a hierarquia correspondente aos tipos da gramática aparecerá

II - Utilizando a Gramática

- 1) Teste o sistema fazendo o parsing da sentença *the cat chased the dog*

- 2) Com o mouse na janela TOP clique no item **Parse** do menu
 - a. Selecione Parse input
 - b. Digite a sentença *the cat chased the dog*
 - c. Clique no botão OK

O sistema irá fazer o parsing da sentença e uma janela aparecerá contendo uma pequena árvore de derivação para a única análise desta sentença

- 3) Clique na árvore de derivação com o botão esquerdo do mouse para abrir um menu que permite visualizar uma versão maior da árvore selecionando **Show enlarged tree** e inspecione os nodos da árvore (opções serão mostradas quando o botão esquerdo do mouse for pressionado)
- 4) Experimente o mecanismo de batch parsing para processar um arquivo contendo várias sentenças de teste
 - a. Vá para o diretório lkb/src/grammar1 e crie um arquivo chamado test.out
 - b. Na janela TOP clique no botão **Parse** e depois no item do menu **Batch parse** ...que irá abrir uma janela pedindo o nome do arquivo a ser processado
 - c. Clique no arquivo test.items no diretório grammar1 e depois pressione o botão OK. Isto irá abrir uma nova janela pedindo pelo nome do arquivo onde os resultados serão armazenados
 - d. Entre o nome do arquivo test.out e clique o botão OK. Na janela do Lisp o sistema irá imprimir a mensagem:
 - i. *Parsing test file* quando inicia o arquivo e
 - ii. *Finished test file* quando termina
 - e. Abra o arquivo test.out e investigue os resultados
 - f. Abra o arquivo lexicon.tdl (com Wordpad) e veja os itens lexicais disponíveis
 - g. Abra o arquivo test.items e adicione sentenças usando os itens lexicais em lexicon.tdl e batch parse o resultado
 - h. Veja que sentenças podem ser parseadas e quais não podem
 - i. Agora analise algumas destas sentenças usando Parse input ...

III Extendendo a Gramática

Nesta seção você continuará extendendo a grammar1.

- 1) Adicione o animal de sua escolha como um substantivo
 - a. No Wordpad abra o arquivo lexicon.tdl
 - b. Copie as 5 linhas do item lexical dog e modifique a sua cópia fazendo com que o valor de ORTH seja apropriado para o animal da sua escolha.
 - c. Salve a versão modificada do arquivo
- 2) Recarregue a gramática para as mudanças serem reconhecidas pelo LKB
 - a. Na janela TOP do LKB selecione **Load/ Reload grammar**
 - b. Faça o parsing da sentença *the cat chased the <your-animal>*
 - c. Adicione a sentence a test.items e faça um Batch parse
- 3) Investigue a gramática para ter uma idéia de como funciona. Em particular analise as seguintes sentenças e tente decidir porque elas parseiam ou não

- a. the cat barks
- b. the cat chased
- c. cat barks
- d. the cat bark
- e. bark

Perguntas:

Se você adicionar estas sentenças em test.items, o que são os números que aparecem ao lado de cada sentença após o **Batch Parsing**? Dica: compare com o resultado de fazer o parsing individual de uma sentença com o **Parse input**

Para fazer isto examine as árvores de derivação (**parse tree**) se estiverem disponíveis e a **parse chart**. Note que a parse chart estará disponível mesmo que a sentence não possa ser parseada.

- 4) Na janela TOP selecione **Parse/Show parse chart** e você poderá ver a parse chart para a sentença.
- 5) A regra que é necessária para capturar ditransitivos está na gramática, mas não há entradas que utilizem a mesma. Adicione uma entrada para **gave** que precise de dois NPs (noun phrases) como complementos (a entrada que seria necessária para parsear **that dog gave the cat the <your_animal>**.
 - a. Copie a entrada de **chased** de lexicon.tdl
 - b. Substitua o valor de ORTH como antes
 - c. Adicione um elemento extra em COMPS (complements - elementos necessarios a direita), que será igual aquela que já está lá. Listas são delimitadas por < e > e os elementos da lista são separados por vírgulas.
 - d. Teste o resultado parseando **that dog gave the cat the <your-animal>**. Teste também para overgeneration confirmando que você não pode parsear **that dog gave the cat**.
 - e. Adicione algumas sentenças para test.items.
- 6) Adicione um novo tipo e duas novas entradas lexicais para parsear **that dog gave the cat to the <your-animal>**
 - a. Adicione o tipo **prep** como um subtipo do tipo **pos** ao arquivo **types.tdl** copiando a descrição para o tipo **noun** e substituindo **noun** por **prep**.
 - b. Adicione uma entrada lexical para a preposição **to**. Ela deve ser similar a entrada de **chased** pois ela tem uma única noun phrase como complemento, mas o valor de HEAD deve ser prep e o valor de SPR (specifier - elementos necessarios a esquerda) tem que ser a lista vazia (representada como <>).
 - c. Adicione uma outra entrada lexical para gave, mas você precisará usar um nome diferente para o identificador (isto é, o string à esquerda de :=), por exemplo, gave_2. Você também precisa trocar o segundo elemento em COMPS para fazer com que esta entrada requeira um PP.

- d. Adicione vários itens de teste, tanto gramaticais quanto não-gramaticais ao arquivo `test.items`, que irão permitir que você cheque que as suas modificações estão corretas.
- e. Rode novamente um batch parse em `test.items` e examine os resultados.