An application of AOC-posets:
Indexing large corpuses
for text generation under constraints
Appendix 1 and 2

Alain Gutierrez, Michel Chein, Marianne Huchard, and Pierre Pompidor

LIRMM, CNRS and Montpellier University, France
{Alain.Gutierrez,marianne.huchard,pierre.pompidor}@lirmm.fr
http://www.lirmm.fr

Abstract. In this paper, we describe the different ingredients of the CogiText tool which can be used for building, editing, and using large corpuses for text generation under constraints *à la* Alamo. In Cogi-Text, AOC-posets are used as indexes that give information about the shape of the corpuses and that help to efficiently find terms for the text creation process. We give some figures about their size and the needed time for computing them and for making a specific text creation. This document contains the appendix 1 and 2 of the ISMIS 2017 submission.

Keywords: Formal Concept Analysis, AOC-poset, Text Generation with Constraints, alamo

Appendix 1: Figures about CogiText corpuses built from DELA and Morphalou

Table 1 shows the figures corresponding to the inflected forms that have been built or computed from DELA and Morphalou.

Table 1. CogiText NOUN corpus, extracted from DELA

DELA nouns	#corpus elements: 160268
key	number of different values
txt	144873
phon	74090
rhyme1	34
rhyme2	609 valeurs
rhyme3	4804
nbsyl	14 valeurs
gender	2 (_feminine, _masculine)
number	2 (_plural, _singular)

DELA adjectives	#corpus elements: 102024
key	number of different values
txt	74676
phon	33471
rhyme1	33
rhyme2	446
rhyme3	2605
nbsyl	13
gender	2 (_feminine, _masculine)
number	2 (_plural, _singular)

Morphalou verbs	#corpus elements: 376851
key	number of different values
txt	285200
phon	146231
rhyme1 (size 1)	31
rhyme2 (size 2)	404
rhyme3 (size 3)	3048
nbsyl	9
number	2 (_plural, _singular)
tense	5 (future, imperfect, ...)
mode	6 (conditional, imperative, ...)
personne	3 (first, second, third)

Appendix 2: Production scheme for "Le corbeau et le renard

```
//Le corbeau et le renard (The crow and the fox)
importScript("attributs/attributs_standard"); // attribute definition
corpusNom=corpus("dela.nom.jss"); // The used noun corpus
corpusAdj=corpus("dela.adjectif.jss"); // The used adjective corpus
/* le corbeau (the crow) */
X1=element(corpusNoun);
eq(X1.rhyme3,"Rbo");
eq(X1.nbsyl,2);
eq(X1.gender,"_masculine");
eq(X1.number,"_singular");
/* le renard (the fox) */
X2=element(corpusNoun);
eq(X2.rhyme3,"naR");
eq(X2.nbsyl,3);
eq(X2.gender,"_masculine");
eq(X2.number,"_singular");
/* l'arbre (the tree) */
X3=element(corpusNoun);
eq(X3.nbsyl,2);
eq(X3.rhyme3,"RbR");
eq(X3.number,"_singular");
/* perché (perched) */
X4=element(corpusAdj);
eq(X4.rhyme3,"RSe");
eq(X4.nbsyl,2);
/* The 'arbre' substitute and the 'perché' substitute
    should agree in number and gender */
eq(X3.gender,X4.gender);
eq(X3.number,X4.number);
eq(X3.number,"_singular");
eq(X4.gender,"_masculine");
/* what has 1 syllable and finishes in 'bEK' */
X5=element(corpusNoun);
eq(X5.rhyme3,"bEk");
eq(X5.nbsyl,1);
eq(X5.number,"_singular");
/* as a 'fromage' */
X6=element(corpusNoun);
eq(X6.rhyme3,"maZ");
eq(X6.nbsyl,3);
eq(X6.number,"_singular");

production(){
    print("Le "+X1.txt+" et le "+X2.txt);
    print("Maitre "+tX1.xt+" sur un "+X3.txt+" "+X4.txt);
    print("Tenait en son "+X5.txt+" un "+X6.txt);
}
```