

# Genic Interaction Extraction Challenge

## Data Format

### 1 File Structure

The LLL format provides a representation of PubMed sentences. It consists of the following fields (one field by line) :

- ID : unique identifier of the Pubmed abstract that contains the sentence and the sentence number
- sentence : the original sentence
- words : sequence of the sentence's words
- agents : list of the agents of the genic interactions
- targets : list of the targets of the genic interactions
- genic\_interactions : list of the interactions described in the sentence.

### 2 Field Structure

A tab separates each element of a field :

#### ID

The ID field contains the abstract's PubMed ID which the sentence is extracted from and it contains the sentence number in this abstract.

ID	(tabulation)	11011148-1
----	--------------	------------

#### SENTENCE

This field contains the sentence.

sentence	(tabulation)	ykuD was transcribed by SigK RNA polymerase from T4 of sporulation.
----------	--------------	---

#### WORDS, AGENTS, TARGETS, GENIC\_INTERACTIONS

Other fields are organised according to the following format :

Field_Name	(tabulation)	predicate1(argument1_1,argument1_2,...)
	(tabulation)	predicate2(argument2_1,argument2_2,...)
	(tabulation)	...

#### EXAMPLE WORDS

words	word(0,'ykuD',0,3)	word(1,'was',5,7)	word(2,'transcribed',9,19)
	word(3,'by',21,22)	word(4,'SigK',24,27)	word(5,'RNA',29,31)
	word(6,'polymerase',33,42)	word(7,'from',44,47)	word(8,'T4',49,50)
	word(9,'of',52,53)	word(10,'sporulation',55,65)	

### 3 Predicate Description

#### WORD

The predicate "word" refers to a word of the sentence and accepts four arguments :

**word(id\_word,'string\_word',char\_word1,char\_word2)**

id_word	integer, unique word id
string_word	string, the actual word
char_word1	integer, position of the first character in the sentence (starting at 0)
char_word2	integer, position of the last character in the sentence (starting at 0)

#### AGENT

The predicate "agent" refers to the agent of the genic interaction. It accepts one argument :  
**agent(id\_word)**

id_word	integer, id of the word the agent refers to
---------	---

#### TARGET

The predicate "target" refers to the target of the genic interaction. It accepts one argument :  
**target(id\_word)**

id_word	integer, id of the word the target refers to
---------	--

#### GENIC\_INTERACTION

The predicate "genic\_interaction" refers to an interaction between an agent and a target :  
**genic\_interaction(id\_word1,id\_word2)**

id_word1	integer, id of the word the agent refers to
id_word2	integer, id of the word the target refers to

### 4 Example

ID	11011148-1
sentence	ykuD was transcribed by SigK RNA polymerase from T4 of sporulation.
words	word(0,'ykuD',0,3) word(1,'was',5,7) word(2,'transcribed',9,19) word(3,'by',21,22) word(4,'SigK',24,27) word(5,'RNA',29,31) word(6,'polymerase',33,42) word(7,'from',44,47) word(8,'T4',49,50) word(9,'of',52,53) word(10,'sporulation',55,65)
agents	agent(4)
targets	target(0)
genic_interactions	genic_interaction(4,0)