

Милинцевич Кирилл
Садов Михаил

База данных повторов в текстах (контекст, в котором производится поиск повторов – одно предложение; под повтором понимается наличие двух или более слов в пределах контекста, имеющих одинаковую основу).

Обозначения в диаграмме:

gram_cat - грамматические признаки слова в формате MULTTEXT

FK- Foreign Key

Нумерация каждого предложения - уникальна, т.е. в новом тексте нумерация предложений продолжается с того места, где закончилась нумерация в предыдущем.

Вся информация хранится в сущности словоформы. Таким образом, мы можем для каждой словоформы проследить из какого она предложения, какая у неё в нём позиция, её основу, и повторяется она или нет.

Связь между сущностями Wordform и Repetition устроена так, что словоформа может содержать id повторения, а может и не содержать, т.е. нули допустимы.

IDEF1-диаграмма:

Wordform

