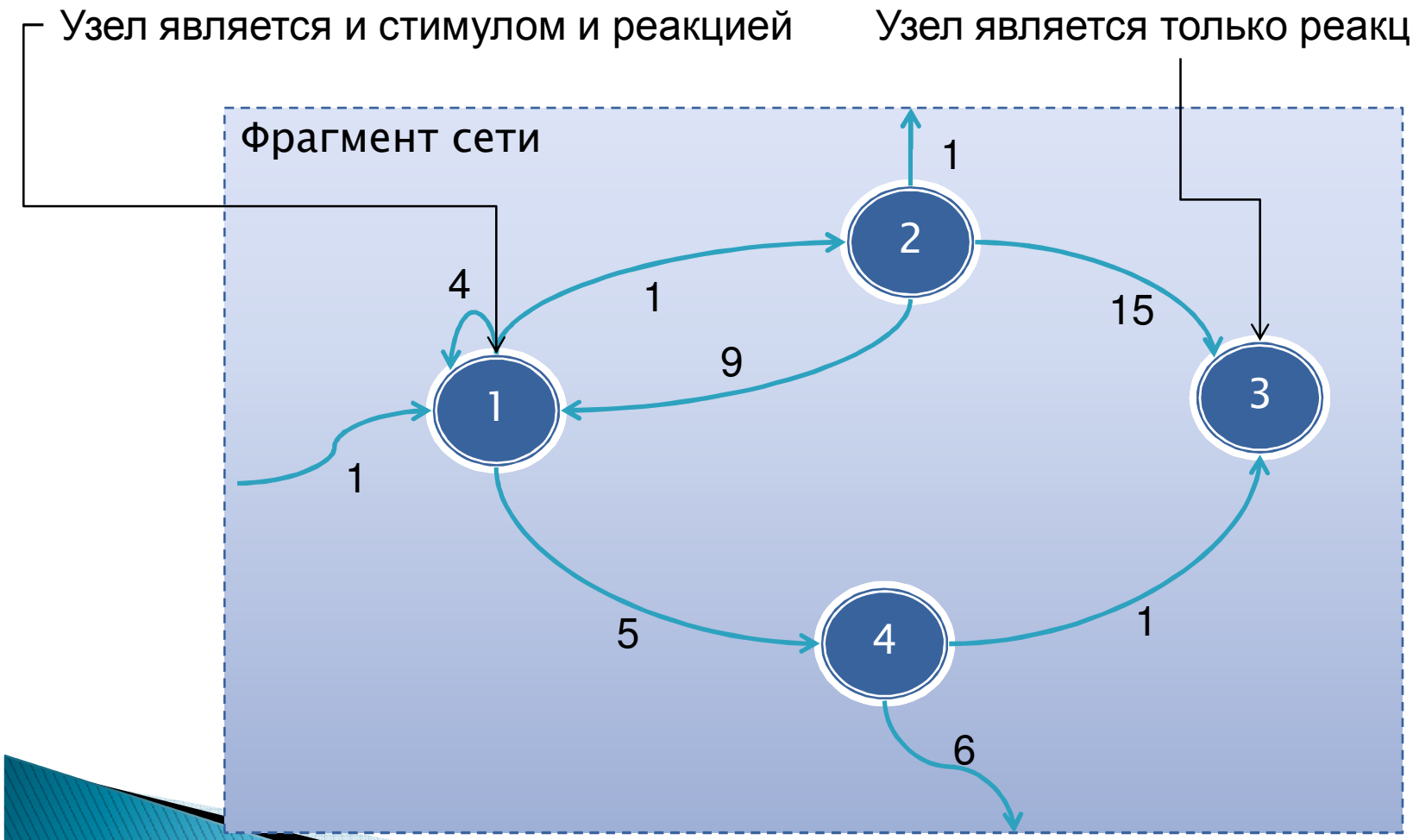
Информационная технология лингвокультурного тезауруса русского языка

Сиренко Александр Викторович
2008

2. Ассоциативная сеть

Число связей: 462 000
Число узлов: 103 000



Число стимулов сети: 6624

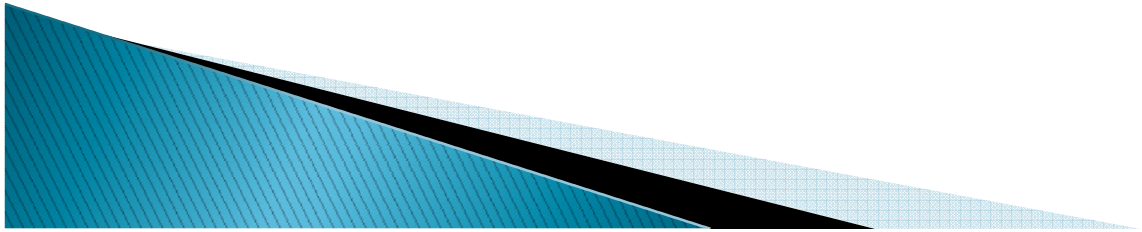
3. Средства разработки

Предъявляемые требования:

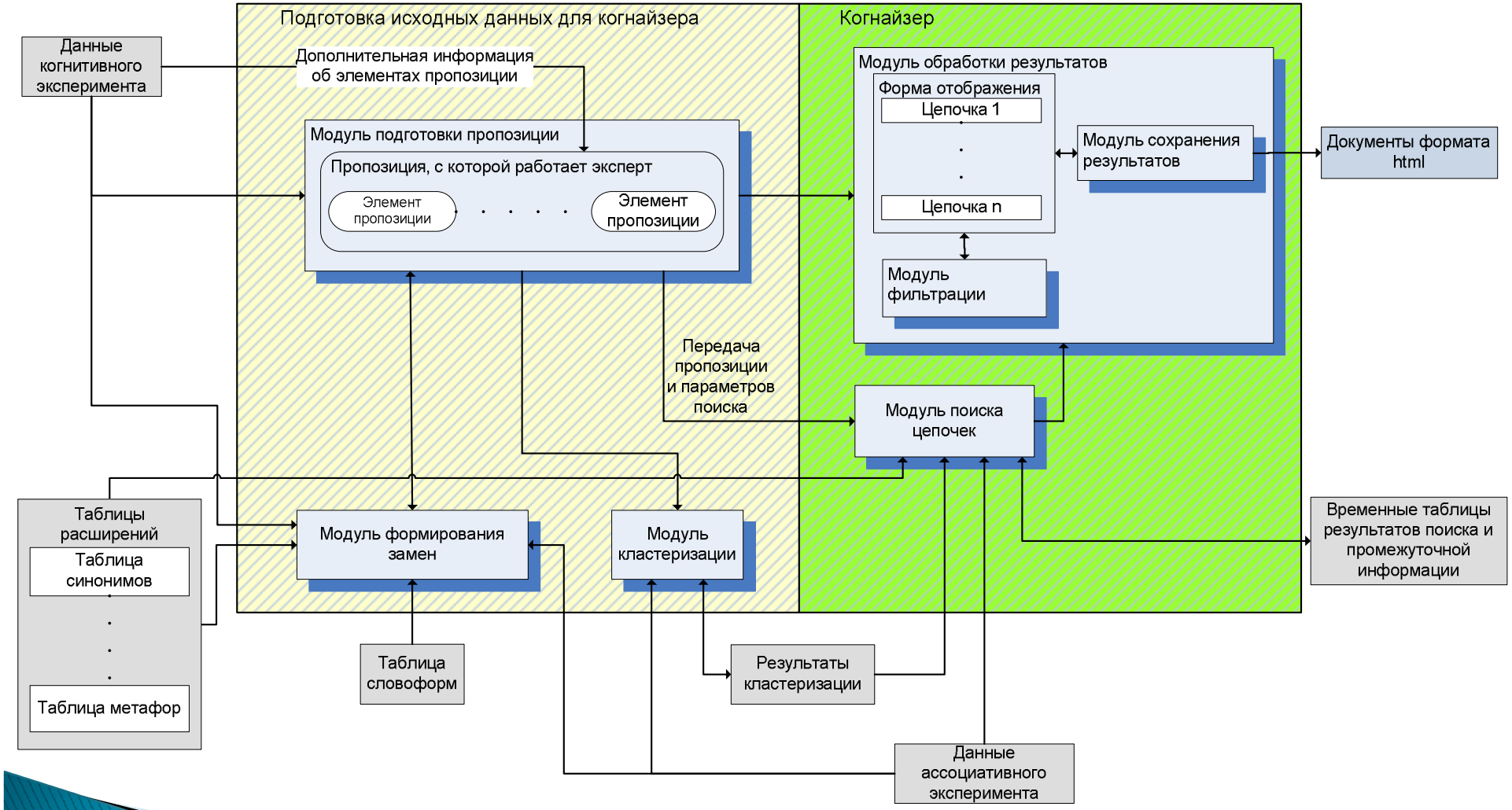
1. Наличие СУБД, не требующей администрирования;
2. Доступность средств разработки, справочной информации, распространенность;
3. Простота установки и обслуживания конечного продукта;
4. Стоимость средств разработки;
5. Отсутствие необходимости покупки лицензий конечным пользователем.

Выбранные средства:

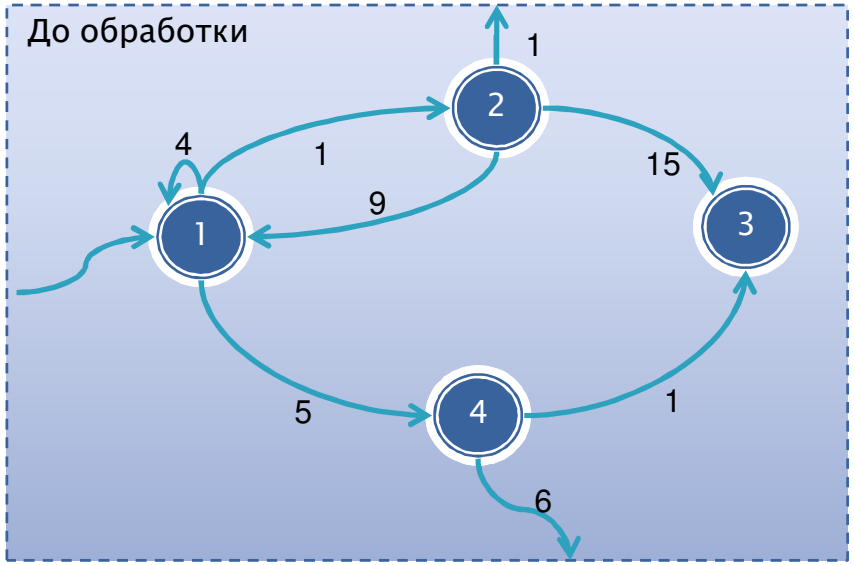
1. Среда разработки: Java SE Developer Kit 6
   http://java.sun.com/
2. Средство проектирования: Sun Microsystems NetBeans 6.5
   http://www.netbeans.org/
3. СУБД: H2
   http://www.h2database.com/



4. Структурная схема лингвокультурного тезауруса русского языка

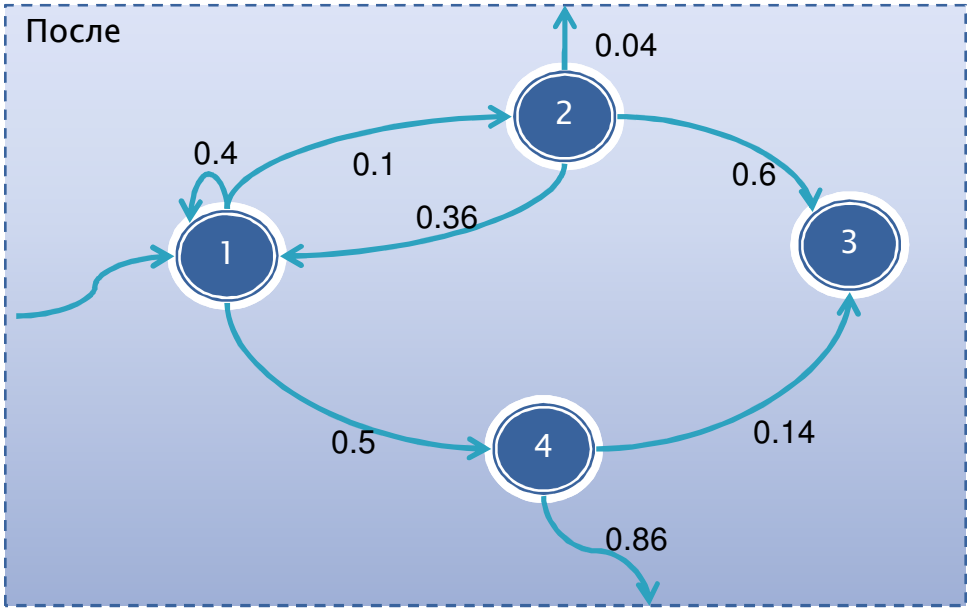


5. Подготовка сети

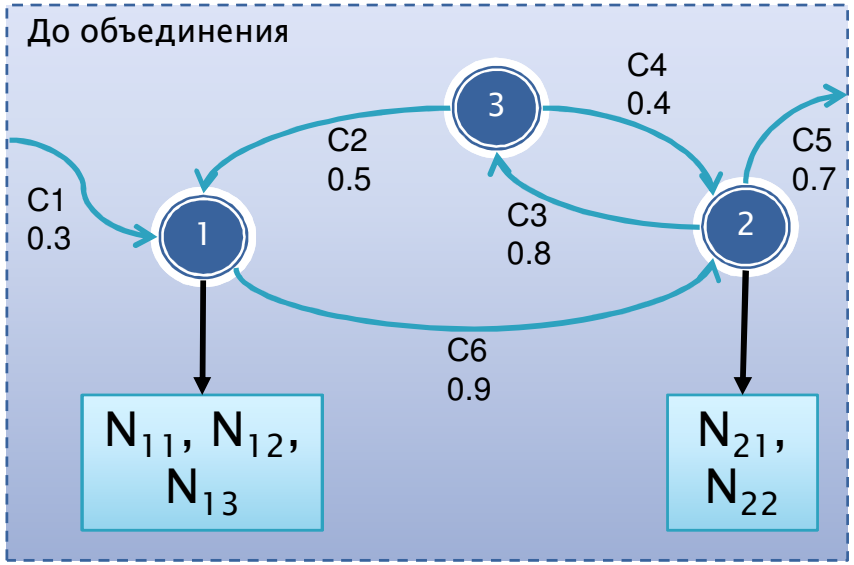


Сумма связей,
выходящих из узла 1 = 4 + 1 + 5 = 10

Делим вес каждой дуги,
выходящей из узла 1 на 10

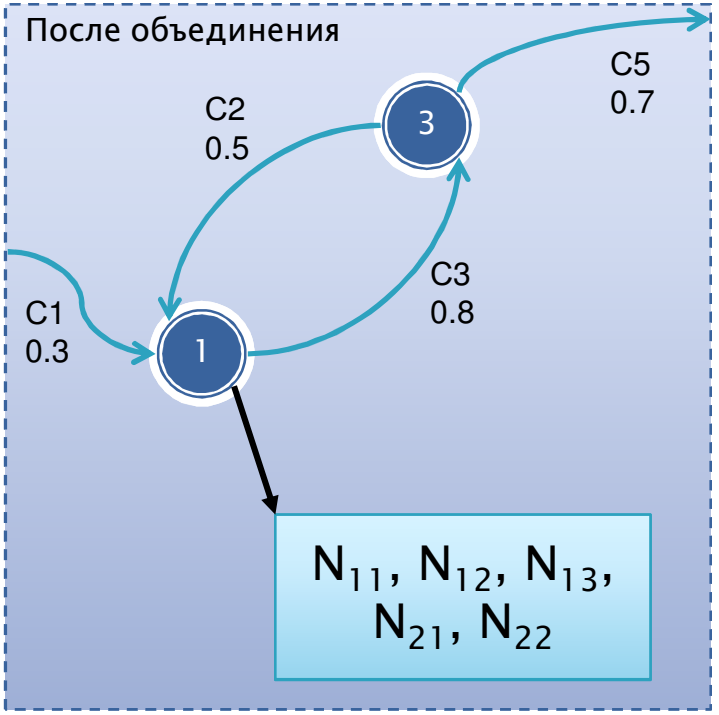


6. Кластеризация

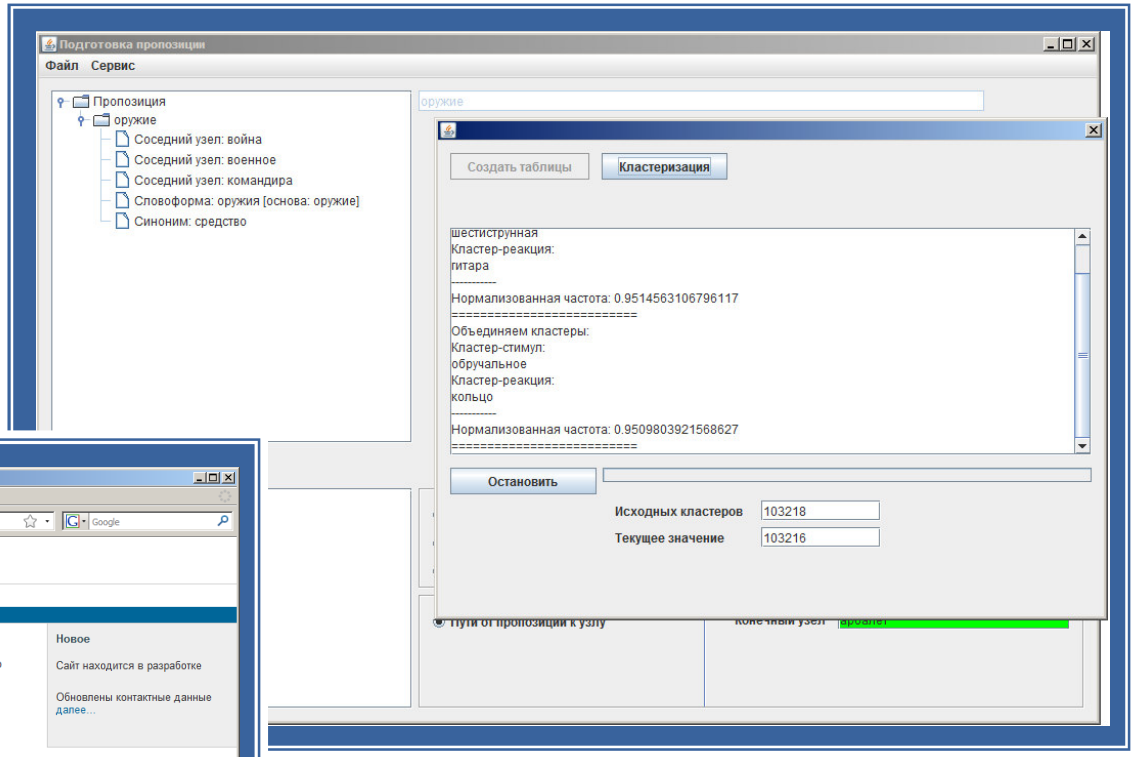


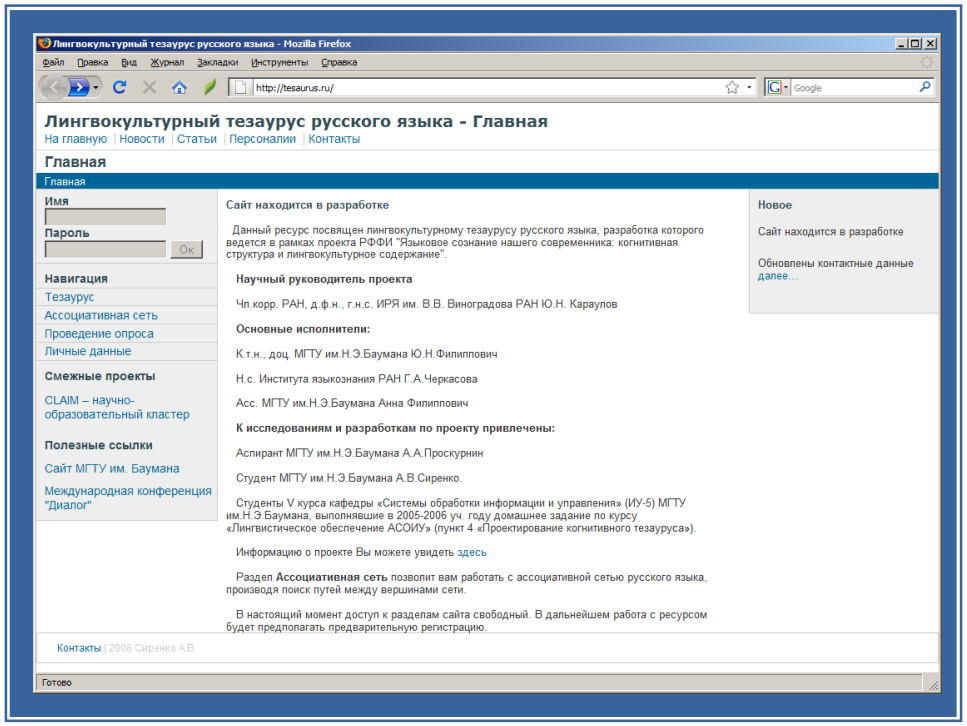
1. Кратчайшая связь C6, следовательно, объединяем кластеры 1 и 2;
2. Переносим узлы N_{21} и N_{22} в кластер 1;
3. Удаляем связь C6;
4. Из связей C2 и C4 оставляем C2;

4. Из связей C2 и C4 оставляем C2;
5. Для связей C3 и C5 указываем кластер 1 в качестве стимула;



7. Заключение





http://tesaurus.ru/