Технологии извлечения знаний из неструктурированных текстов

Разработка практических заданий и лабораторных работ для дисциплин блока "Компьютерная лингвистика»

Карпов Николай
Бонч-Осмоловская Анастасия
Сибирцева Вера
Савченко Андрей
Малафеев Алексей

Ontos
NOW YOU KNOW

Постановка задачи

Текст **Модель** **Структурированные данные**

МОСКВА, 15 мая - РИА Новости. Руководитель Росатома Сергей Кириенко 19-23 мая в ходе поездки в США проведет ряд рабочих встреч, посвященных двустороннему сотрудничеству в области мирного использования атомной энергии, говорится в сообщении пресс-службы Росатома. Планируется, что Кириенко 22 мая проведет переговоры с министром энергетики США Самюэлом Бодманом и руководителем комиссии по ядерному регулированию США Нильсом Диазом.

типы объектов и типы отношений

РАБОТАТЬ В ОРГАНИЗАЦИИ

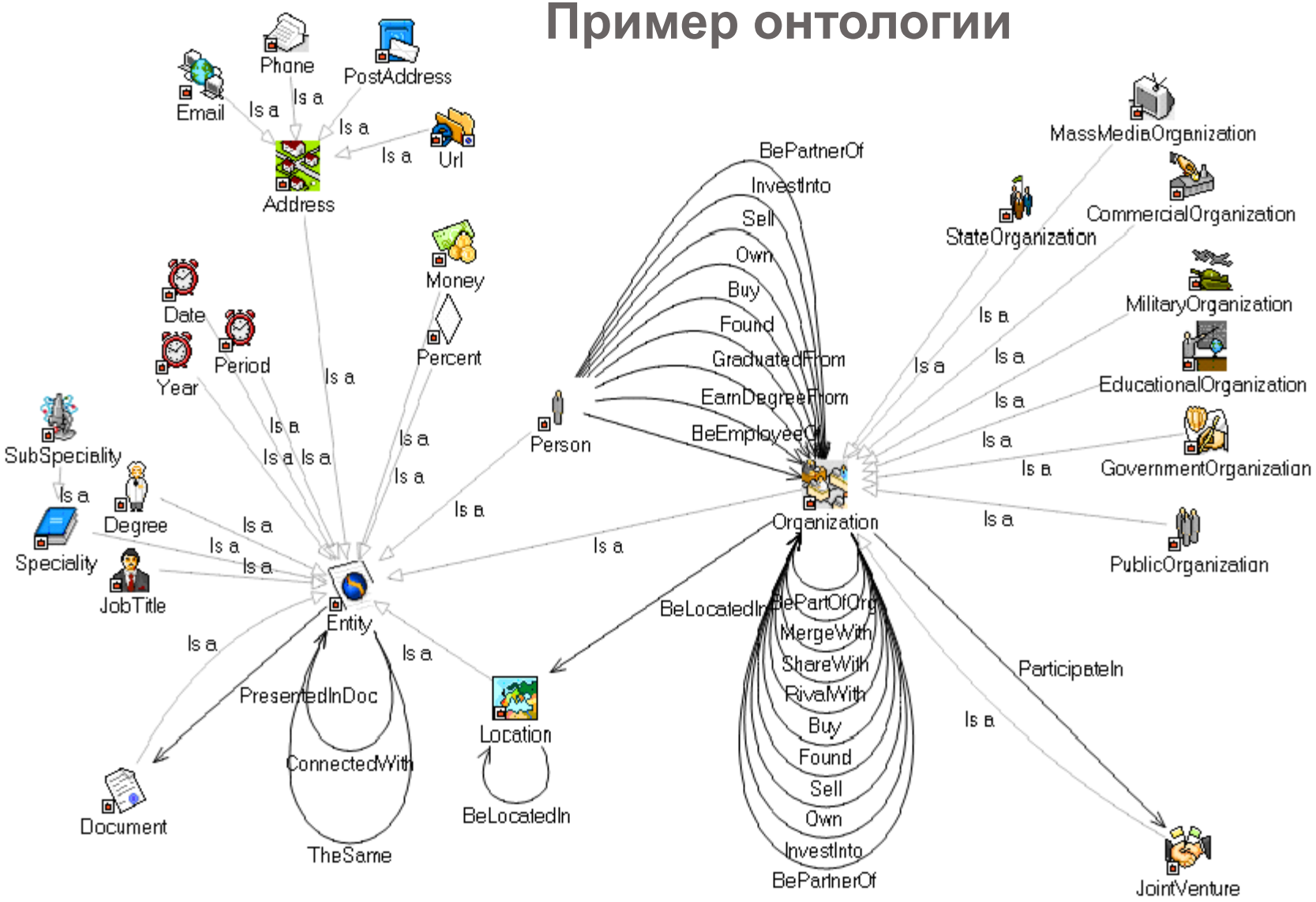
ОРГАНИЗАЦИЯ

ПЕРСОНА

Руководитель Росатома Сергей Кириенко 19-23 мая в ходе поездки в США проведет ряд рабочих встреч...

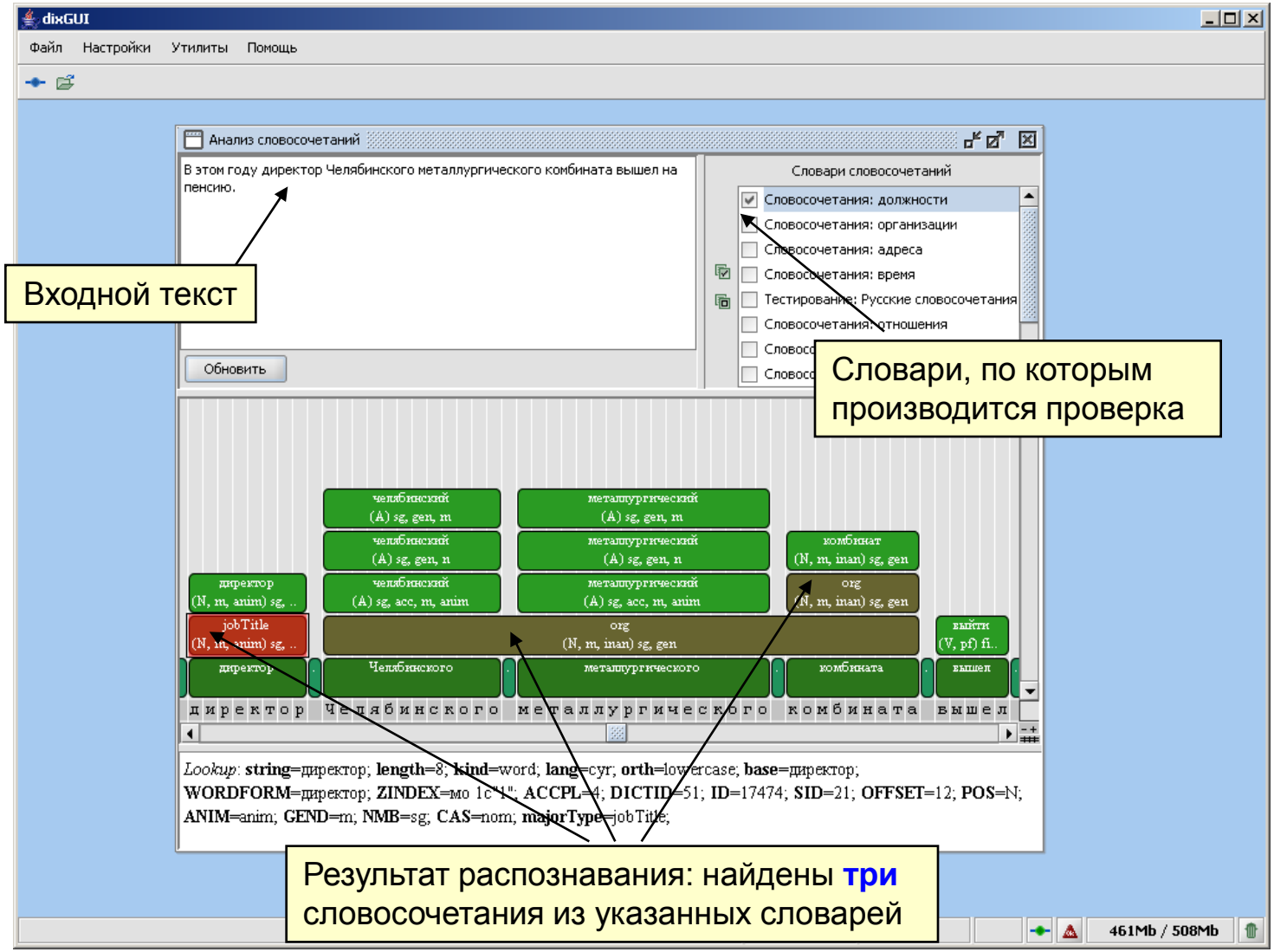
Ontos
NOW YOU KNOW

Пример онтологии



Ontos
NOW YOU KNOW

Распознавание словосочетаний



Ontos
NOW YOU KNOW

Выделение отношений

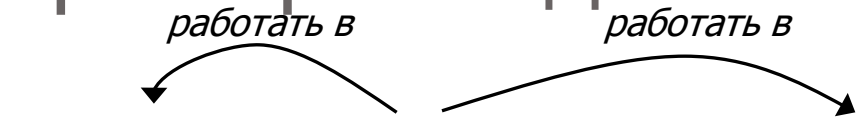
- «Осыпание» нерелевантных фрагментов текста
- Пример с отношением «работать»:

```
({Date} | {StartPoint})? {Person}
{becomeVG.VOICE="act", becomeVG.MOOD="ind"}
{JobTitle} {Organization }
```

В середине 2003-2004 года бывший немецкий гражданин Хайнц Шиммельбуш неожиданно становится директором малоизвестной компании «Васюки».

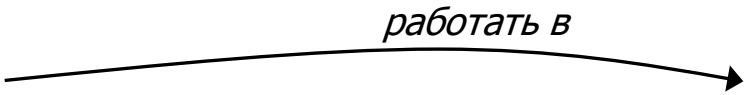
Ontos
NOW YOU KNOW

Примеры выделения отношений

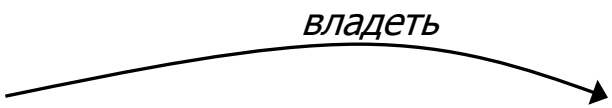


До IBM, г-н. Торощин работал в Microsoft.

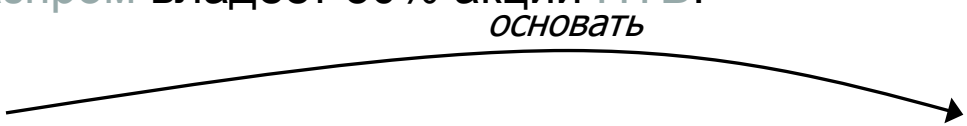
Иванов был уволен из МВД в 1985 году.



И.И. Иванов был принят на работу в РАО ЕЭС.



Компания Газпром владеет 50% акций НТВ.



Мистер Брин был одним из основателей компании Google.

Вторичное распознавание и связывание кореферентных наименований объектов

Иной путь был у нефтяной компании «Сибнефть». Она была образована в 1995 г. на основе ряда предприятий советской нефтяной промышленности. В течение нескольких лет Борис Березовский с Романом Абрамовичем скупили на приватизационных конкурсах контрольный пакет акций компании. По данным Счетной палаты, при продаже компании государству был нанесен ущерб в размере 2,7 млрд. долларов. Об этом заявил глава СП Сергей Степашин.

К «черному золоту» будущий владелец «Сибнефти» имел опосредованное отношение. Учась в институте, Р. Абрамович создал кооператив «Уют», изготавливающий игрушки из полимеров. Лишь в середине 90-х он занялся торговлей нефтью через швейцарскую компанию RUNICOM.

Среди финансовых аналитиков «Сибнефть» при Абрамовиче считалась крайне эффективной компанией.

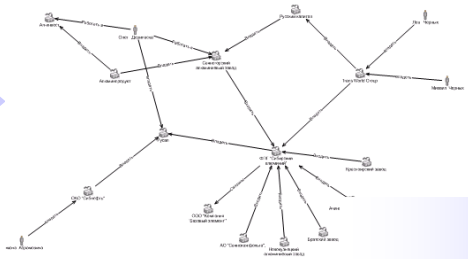
Ontos
NOW YOU KNOW

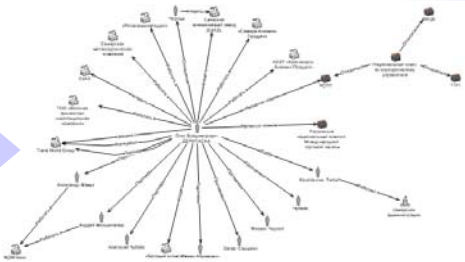
Интеграция информации

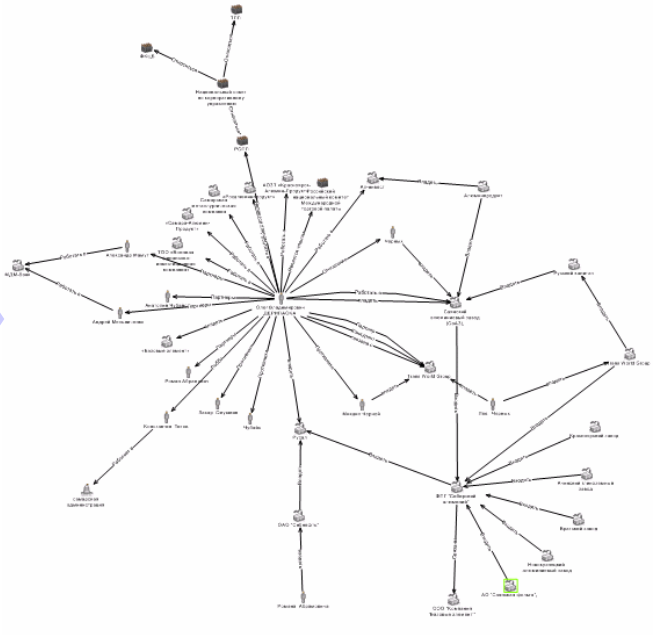
После сохранения необходимо выявить и объединить идентичные объекты из разных документов.

МОСКВА, 15 мая - РИА Новости. Руководитель Росатома Сергей Кириенко 19-23 мая в ходе поездки в США проведет ряд рабочих встреч, посвященных двустороннему сотрудничеству в области мирного использования атомной энергии, говорится в сообщении пресс-службы Росатома. Планируется, что Кириенко 22 мая проведет переговоры с министром энергетики США Самюэлом Бодманом и руководителем комиссии по ядерному регулированию США Нильсом Диазом.

МОСКВА, 15 мая - РИА Новости. Руководитель Росатома Сергей Кириенко 19-23 мая в ходе поездки в США проведет ряд рабочих встреч, посвященных двустороннему сотрудничеству в области мирного использования атомной энергии, говорится в сообщении пресс-службы Росатома. Планируется, что Кириенко 22 мая проведет переговоры с министром энергетики США Самюэлом Бодманом и руководителем комиссии по ядерному регулированию США Нильсом Диазом.







тексты

отдельные графы

база знаний