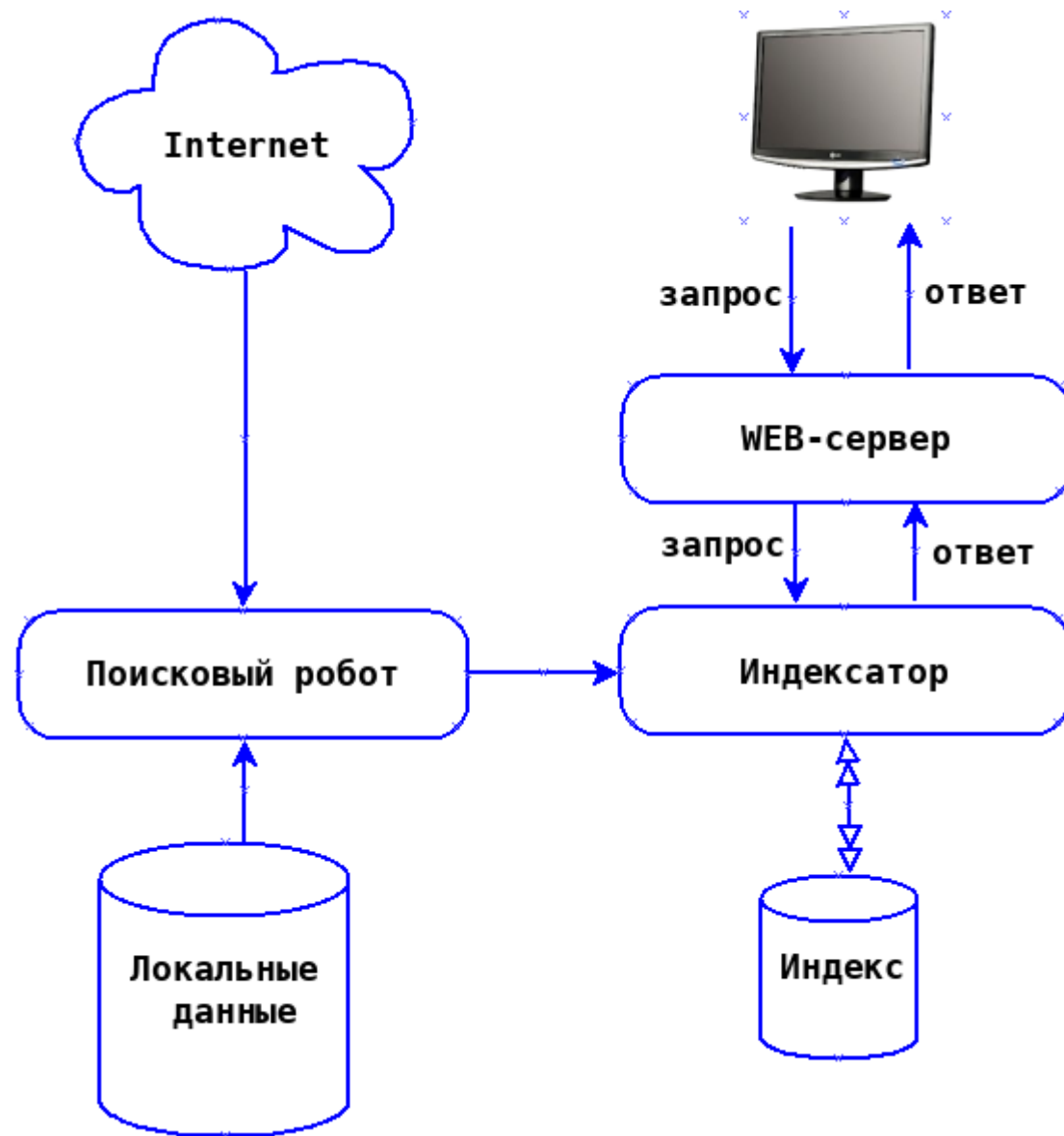


**Применение  
онтологического  
подхода для анализа  
текстов в облачном  
контент-репозитории C2R**

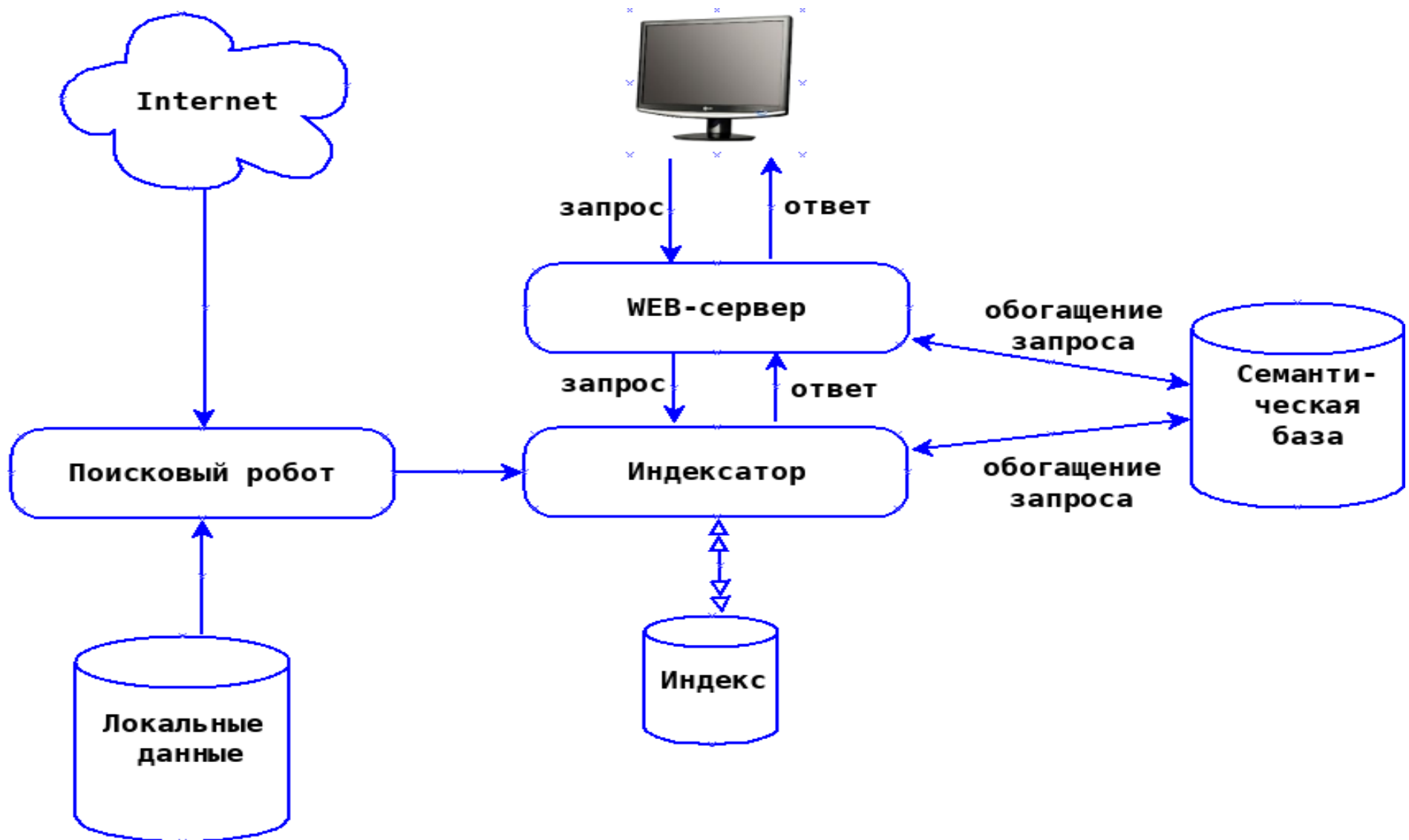
# Участники

- ПГНИУ (Пермский Государственный Научно-Исследовательский Университет)
  - Теоретическая база
  - Редактор онтологий Онтолиз
  - Программирование клиентских и серверных приложений
- ГК ИВС (ИВС, НЕВОД)
  - Репозиторий C2R (Apache Hadoop CDH4, Lily, Solr)
  - Облачный кластер (5 серверов)
  - Программирование клиентских и серверных приложений

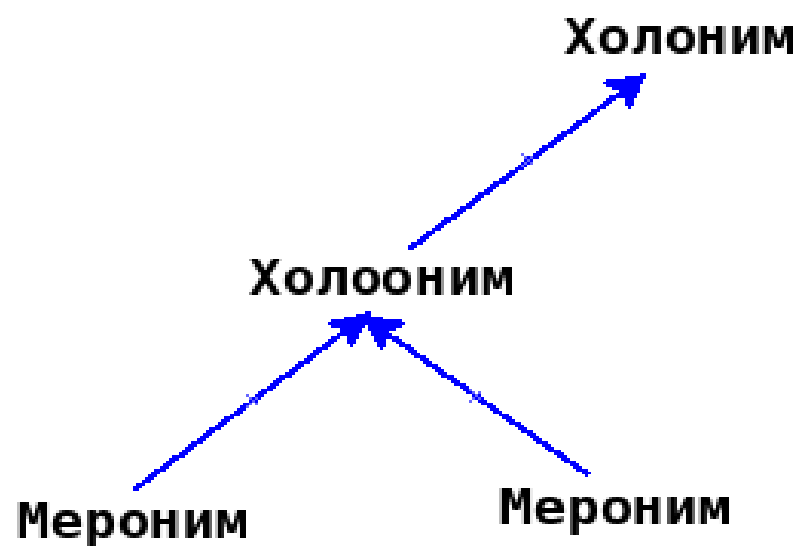
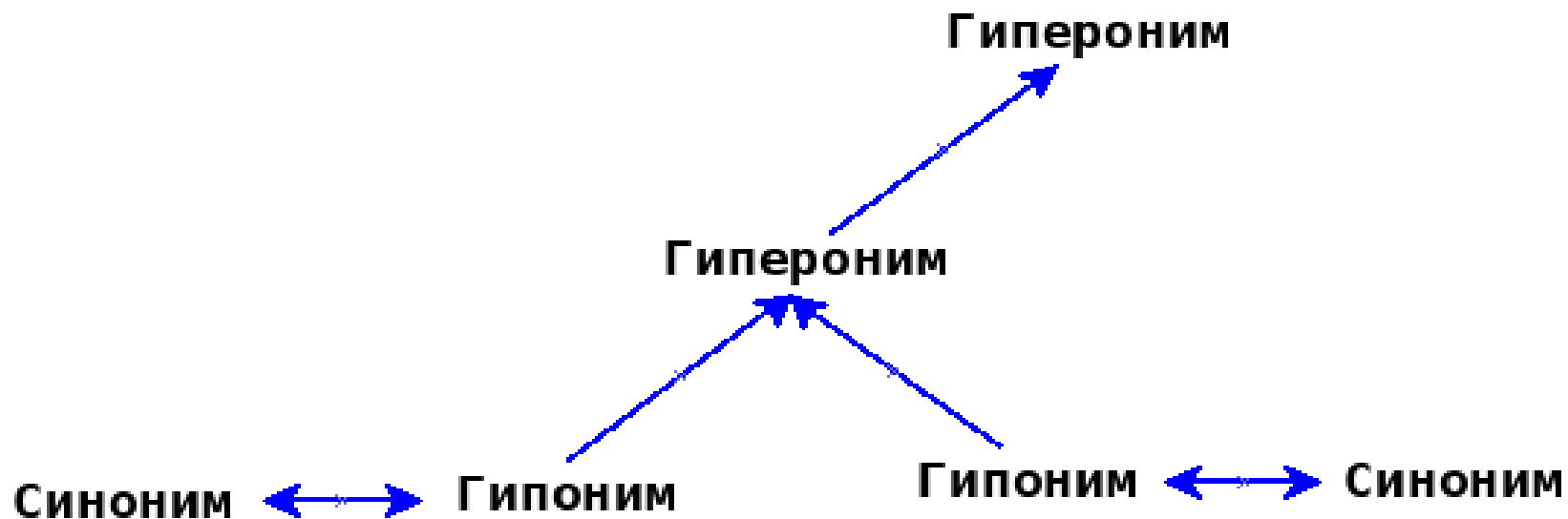
# Простые поисковые системы



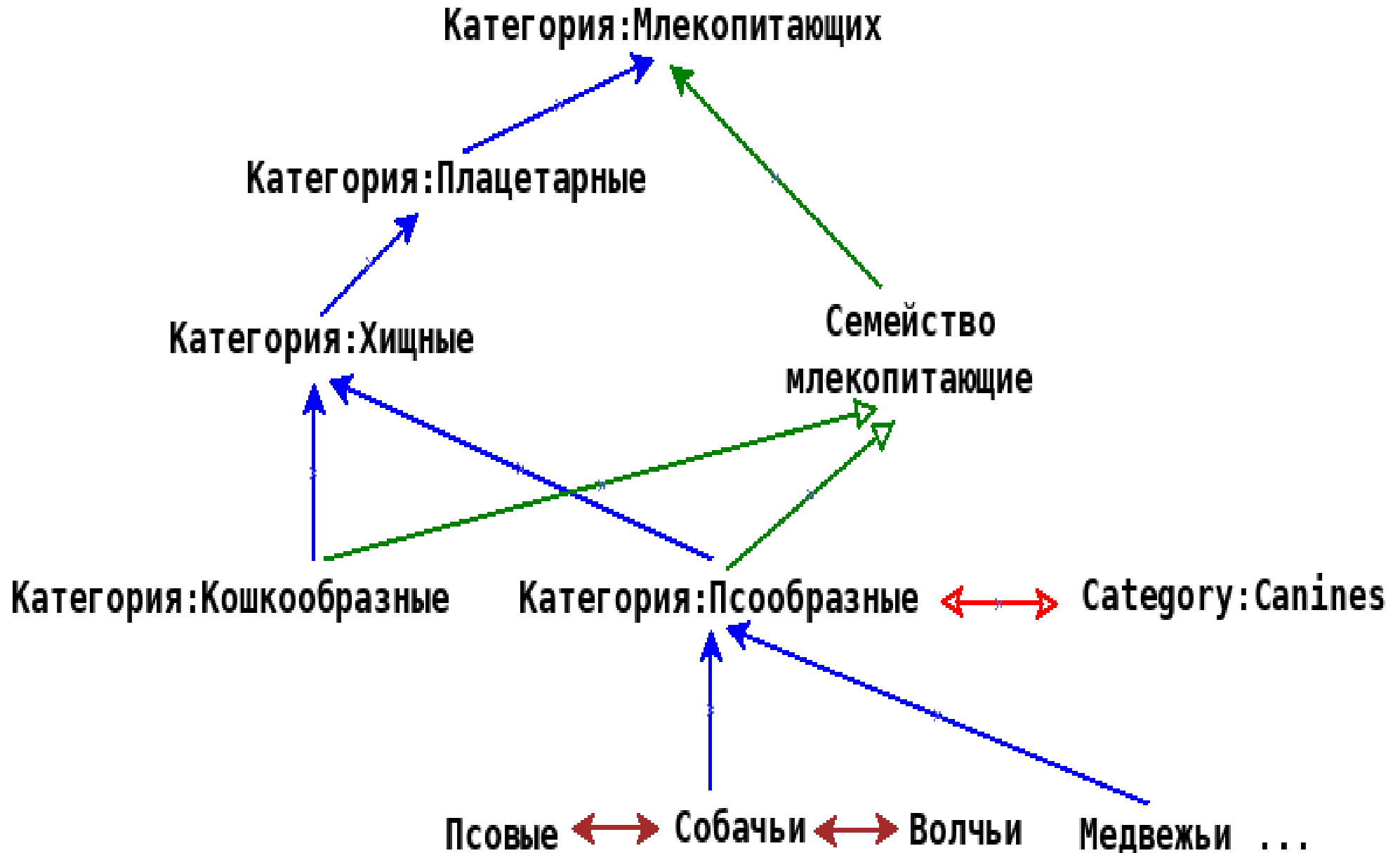
# Семантический поиск



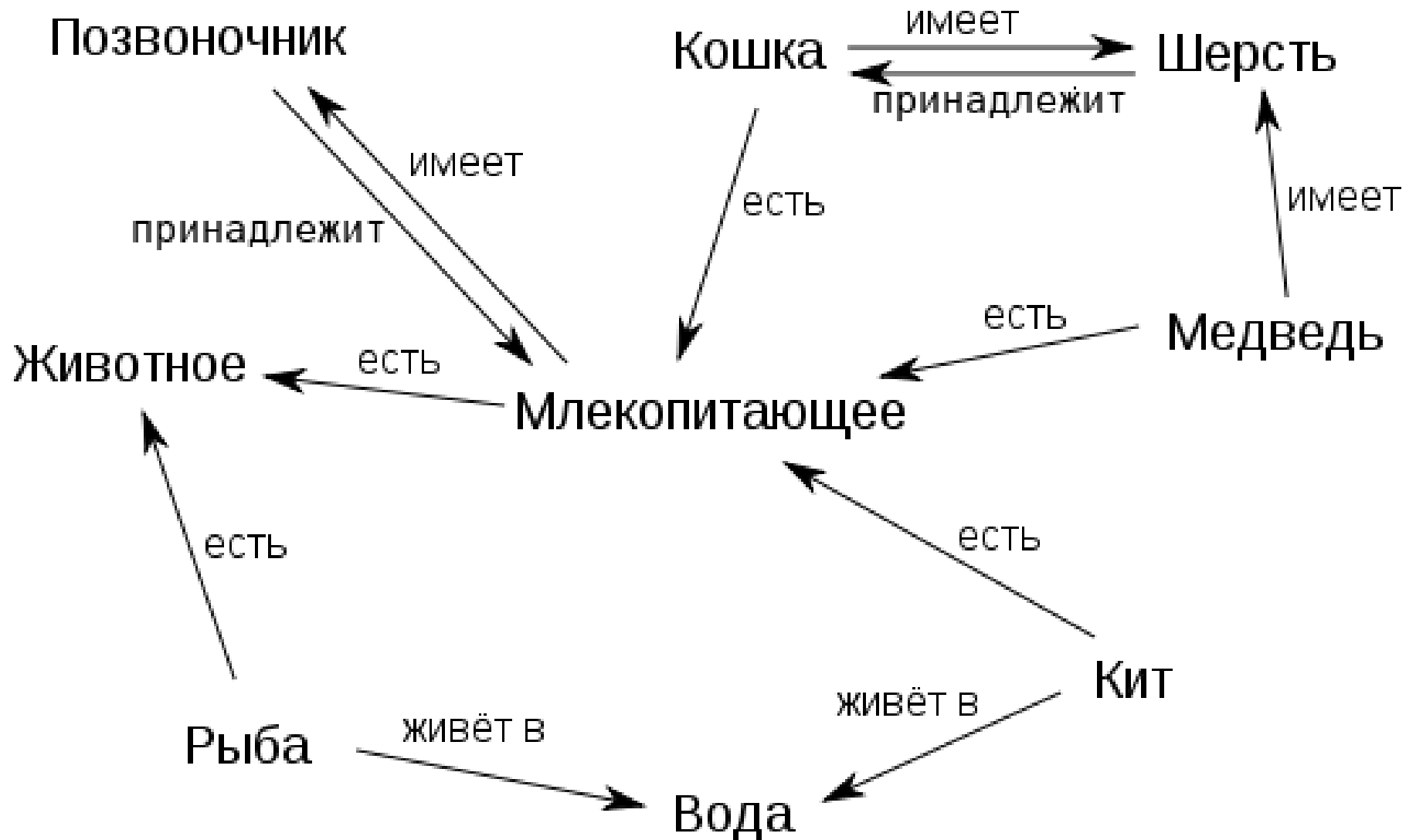
# Семантические отношения



# Пример отношений



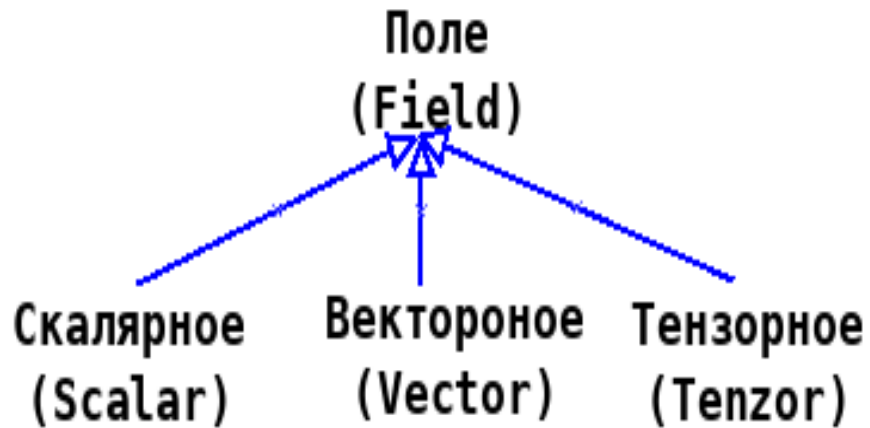
# Семантическая сеть (OWL)



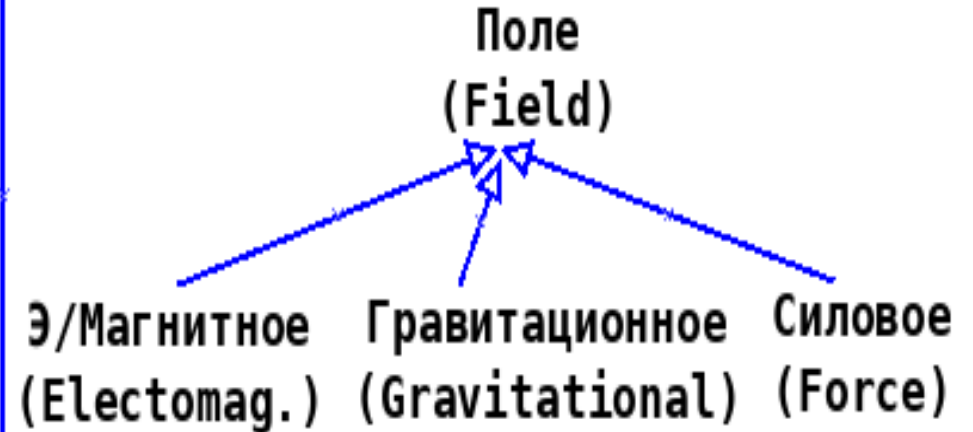
**принадлежит** <-> **имеет** - симметричные отношения  
**есть**, **гипероним**, **гипоним** - транзитивные отношение  
**синоним**, **когипоним** - рефлексивные отношения

# Предметные области: термин «Поле»

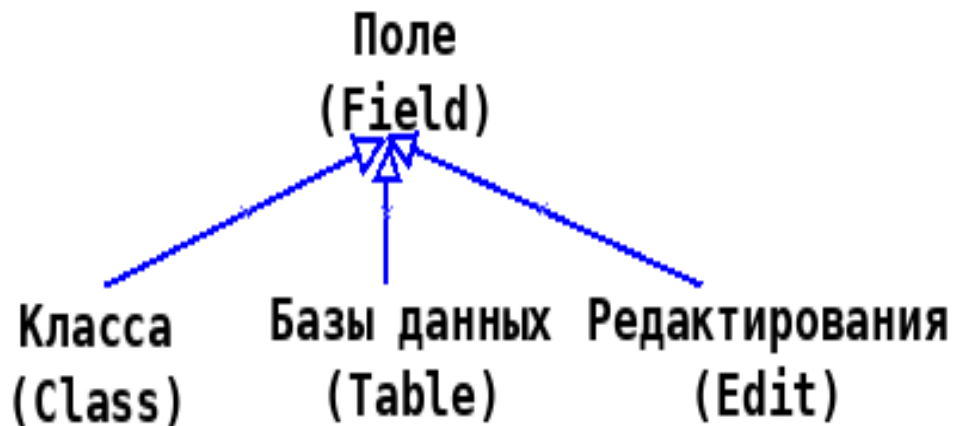
Предметная область: Математика



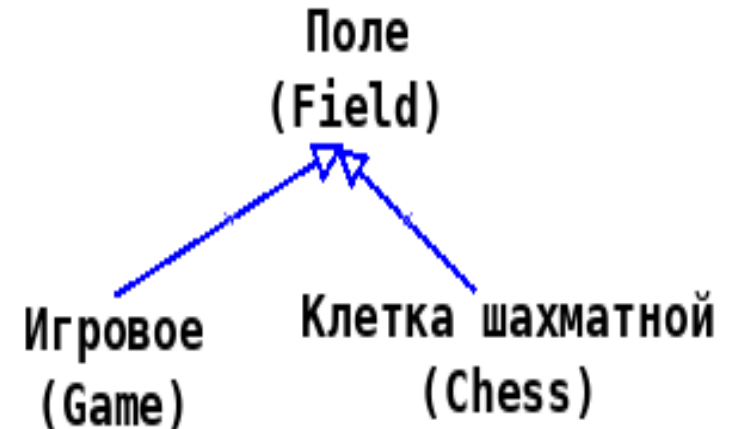
Предметная область: Физика



Предметная область: Информатика



Предметная область: Спорт





# RDF-представление (quad)

<u>Subject</u>	<u>Property</u>	<u>Object(Literal)</u>	<u>Scope</u>
<i>Поле</i>	<u><i>Disambiguates</i></u>	<i>Поле_(Математика)</i>	.
<i>_"_"</i>	<i>_"_"</i>	<i>Поле_(Физика)</i>	.
<i>_"_"</i>	<i>_"_"</i>	<i>Поле_(Информатика)</i>	.
<i>_"_"</i>	<i>_"_"</i>	<i>Поле_(Спорт)</i>	.
<i>Поле_(Математика)</i>	<u><i>SameAs</i></u>	<u><i>Field_(Mathematics)</i></u>	<i>Математика</i>
<i>Поле_(Физика)</i>	<u><i>SameAs</i></u>	<u><i>Field_(Physics)</i></u>	<i>Физика</i>
<i>Поле_(Информатика)</i>	<u><i>SameAs</i></u>	<u><i>Field_(Informatics)</i></u>	<i>Информатика</i>
<i>Поле_(Спорт)</i>	<u><i>SameAs</i></u>	<u><i>Field_(Sport)</i></u>	<i>Спорт</i>
<i>Скалярное поле</i>	<u><i>Subject</i></u>	<i>Поле_(Математика)</i>	<i>Математика</i>
<i>Векторное поле</i>	<u><i>Subject</i></u>	<i>Поле_(Математика)</i>	<i>Математика</i>
<i>Тензорное поле</i>	<u><i>Subject</i></u>	<i>Поле_(Математика)</i>	<i>Математика</i>
<i>Электромагнитное поле</i>	<u><i>Subject</i></u>	<i>Поле_(Физика)</i>	<i>Физика</i>
<i>Гравитационное поле</i>	<u><i>Subject</i></u>	<i>Поле_(Физика)</i>	<i>Физика</i>
<i>Силовое поле</i>	<u><i>Subject</i></u>	<i>Поле_(Физика)</i>	<i>Физика</i>
<i>Поле класса</i>	<u><i>Subject</i></u>	<i>Поле_(Информатика)</i>	<i>Информатика</i>
<i>Поле базы данных</i>	<u><i>Subject</i></u>	<i>Поле_(Информатика)</i>	<i>Информатика</i>
<i>Поле редактирования</i>	<u><i>Subject</i></u>	<i>Поле_(Информатика)</i>	<i>Информатика</i>
<i>Игровое поле</i>	<u><i>Subject</i></u>	<i>Поле_(Спорт)</i>	<i>Спорт</i>
<i>Поле шахматное доски</i>	<u><i>Subject</i></u>	<i>Поле_(Спорт)</i>	<i>Спорт</i>

# Serelex (Семантические метрики)



For example, [local speciality](#)

Show as

Secondary links

Icons

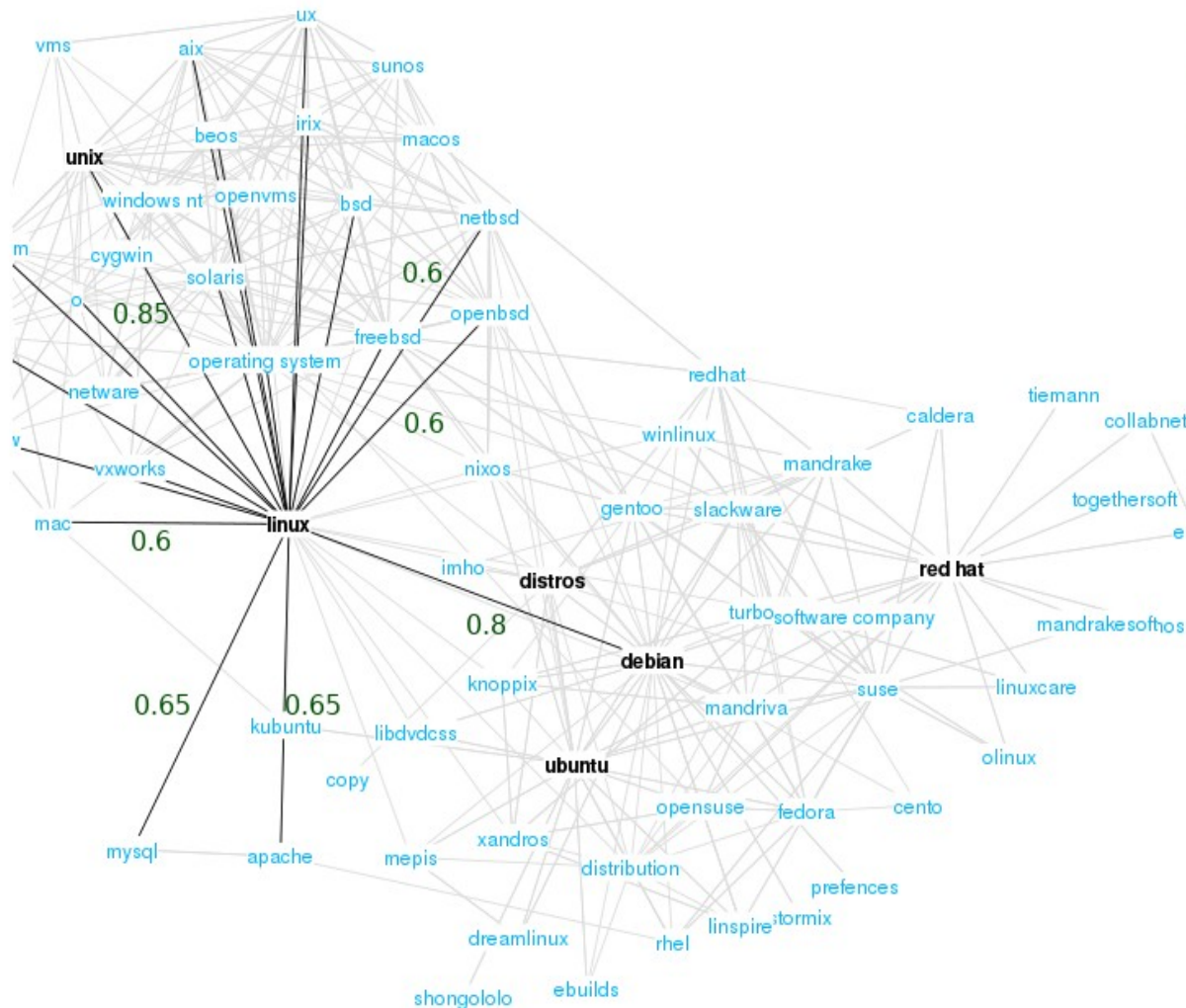
Info panel

Search Language

Results count: 717

- 1 [operating system](#)
- 2 [unix](#)
- 3 [solaris](#)
- 4 [freebsd](#)
- 5 [bsd](#)
- 6 [platform](#)
- 7 [window](#)
- 8 [aix](#)
- 9 [o](#)
- 10 [netbsd](#)
- 11 [ux](#)
- 12 [irix](#)
- 13 [vxworks](#)
- 14 [macintosh](#)
- 15 [apache](#)
- 16 [openbsd](#)
- 17 [mac](#)
- 18 [debian](#)
- 19 [beos](#)
- 20 [mysql](#)

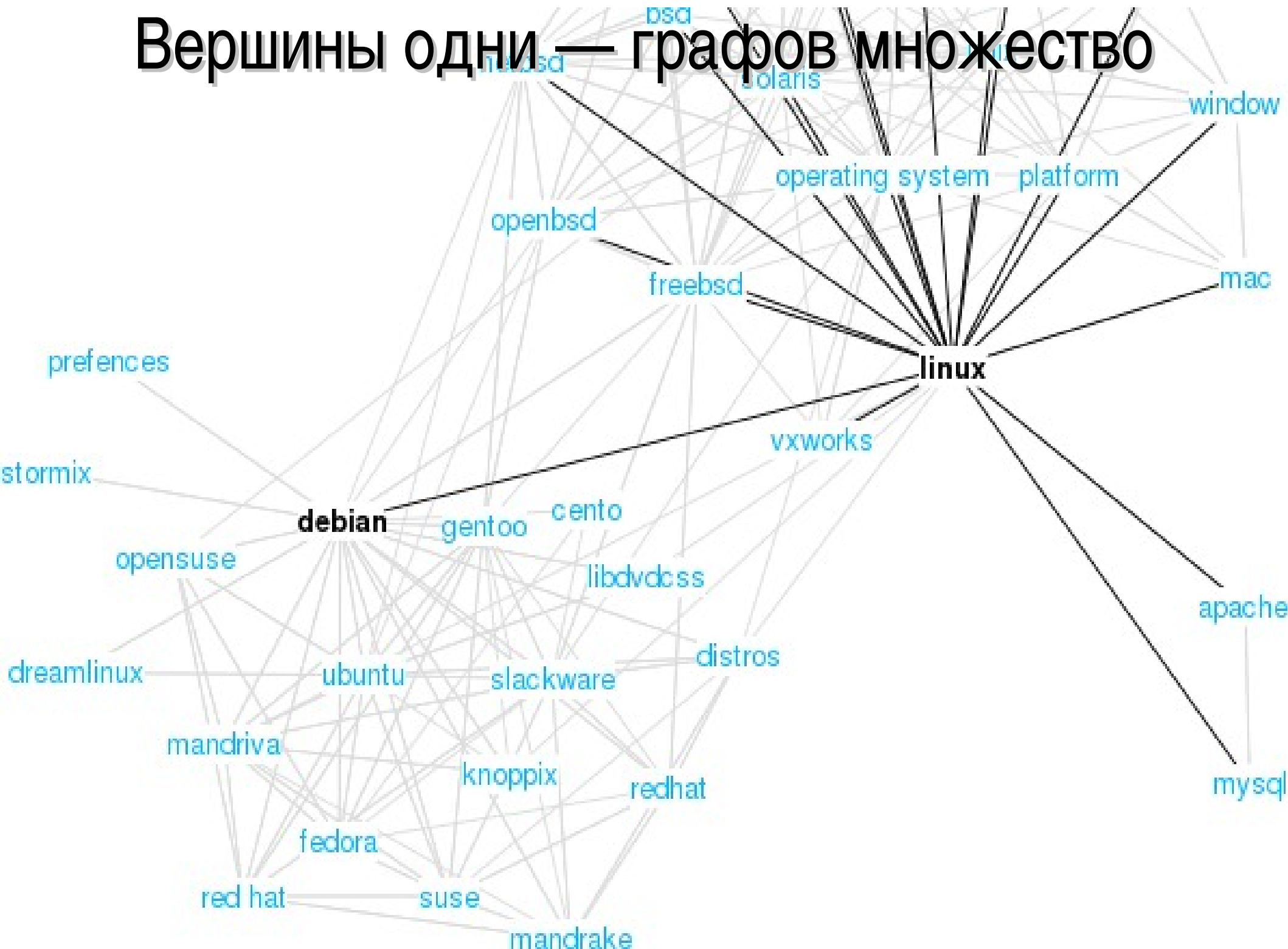
[Show next 20 results](#)

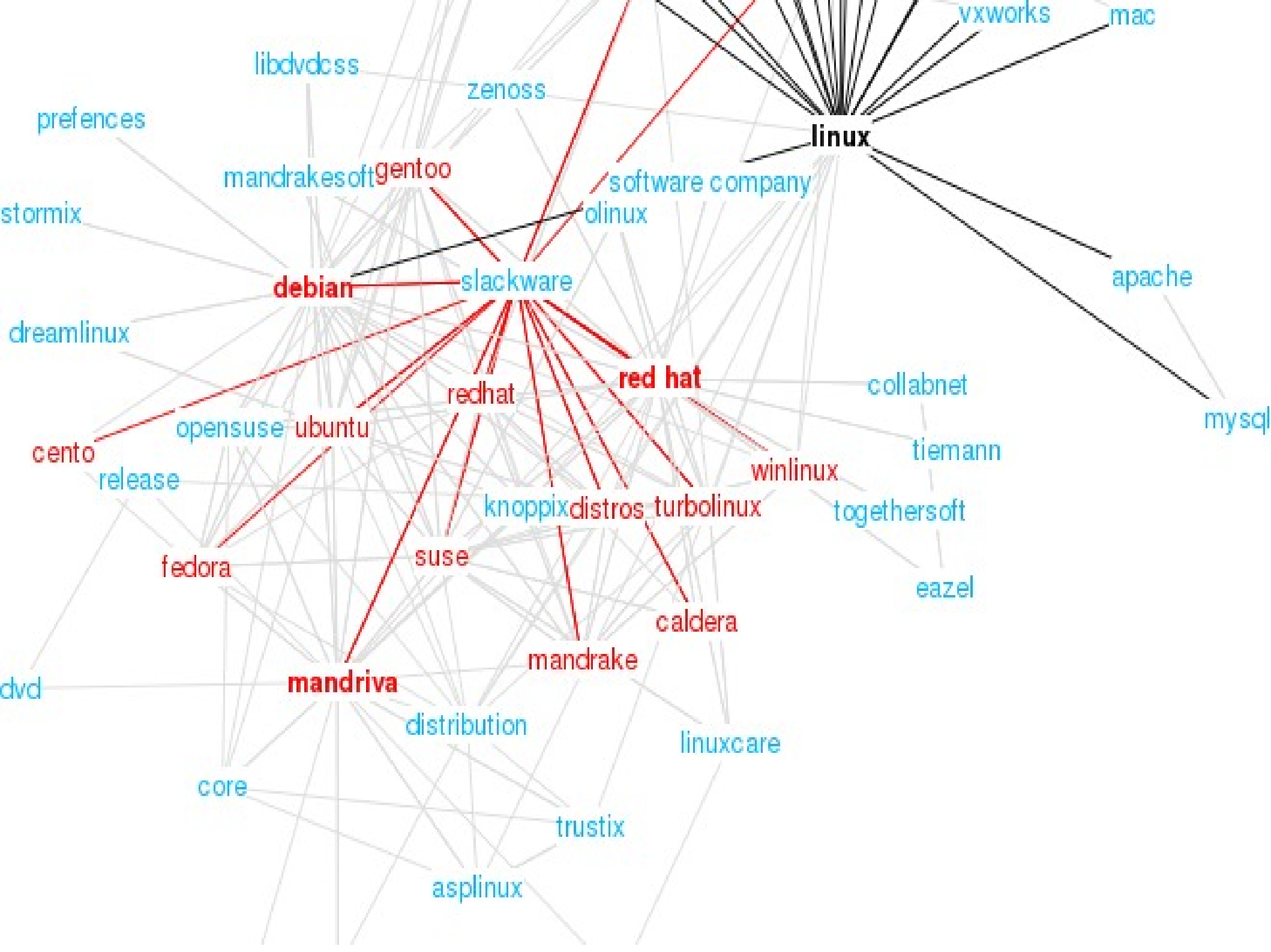


Search in...

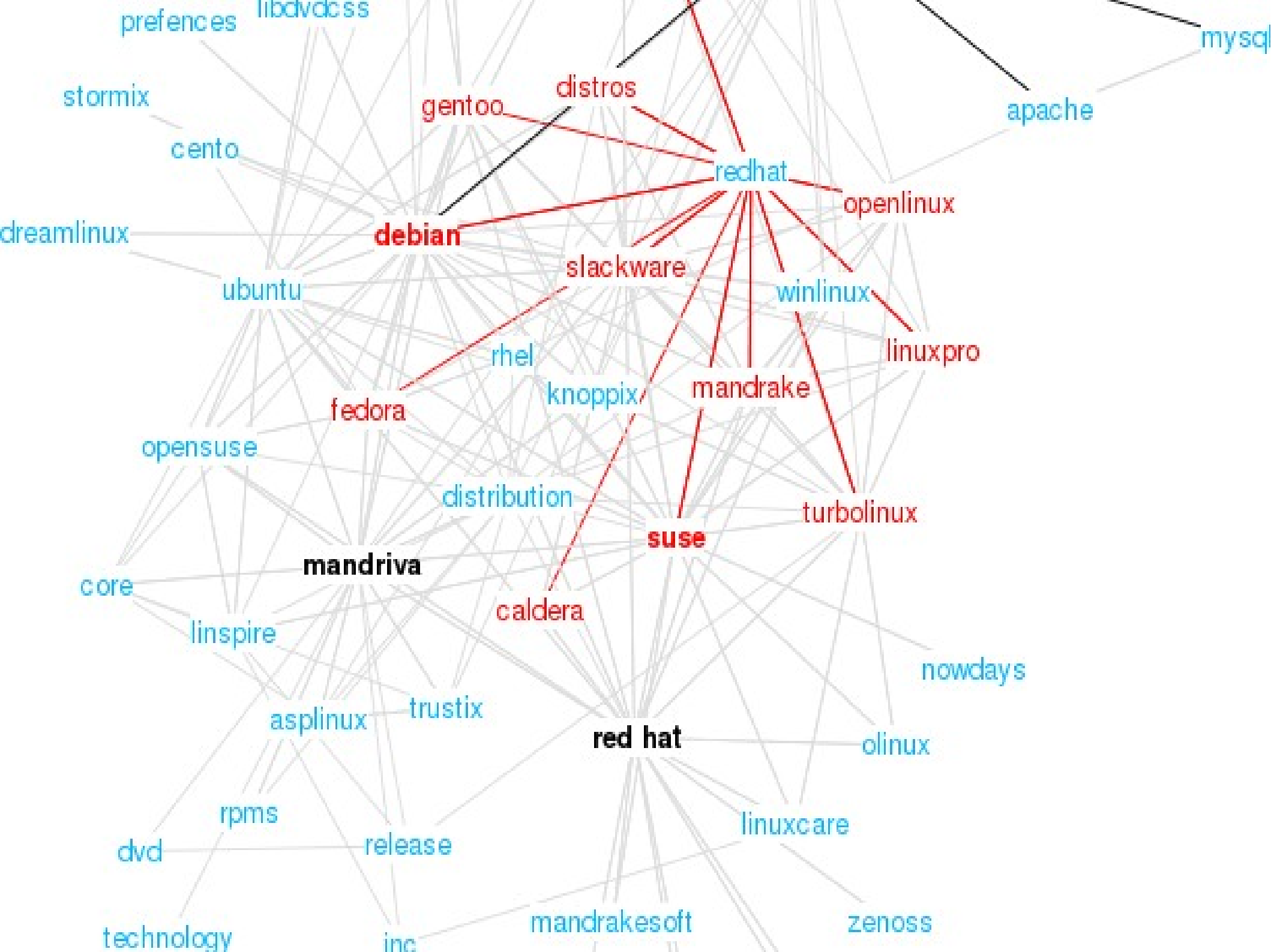
- [Google](#)
- [Thesaurus.com](#)
- [Longman Dictionary](#)
- [The Free Dictionary](#)
- [Oxford Dictionaries](#)
- [Merriam-Webster](#)
- [Cambridge Dictionary](#)
- [Collins Dictionary](#)

# Вершины одни — графов множество





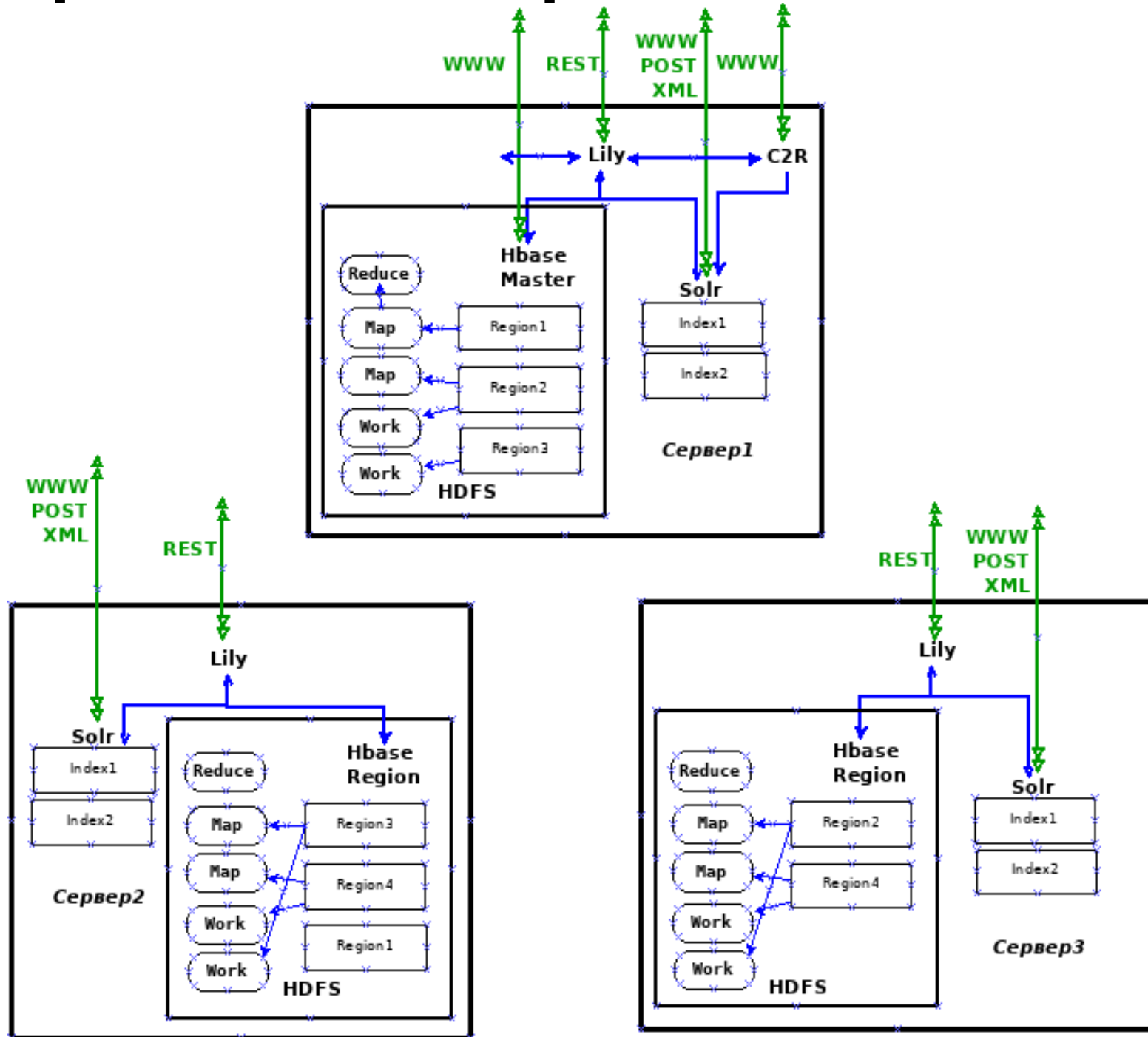




# Представление RDF в C2R/Lily/HBase

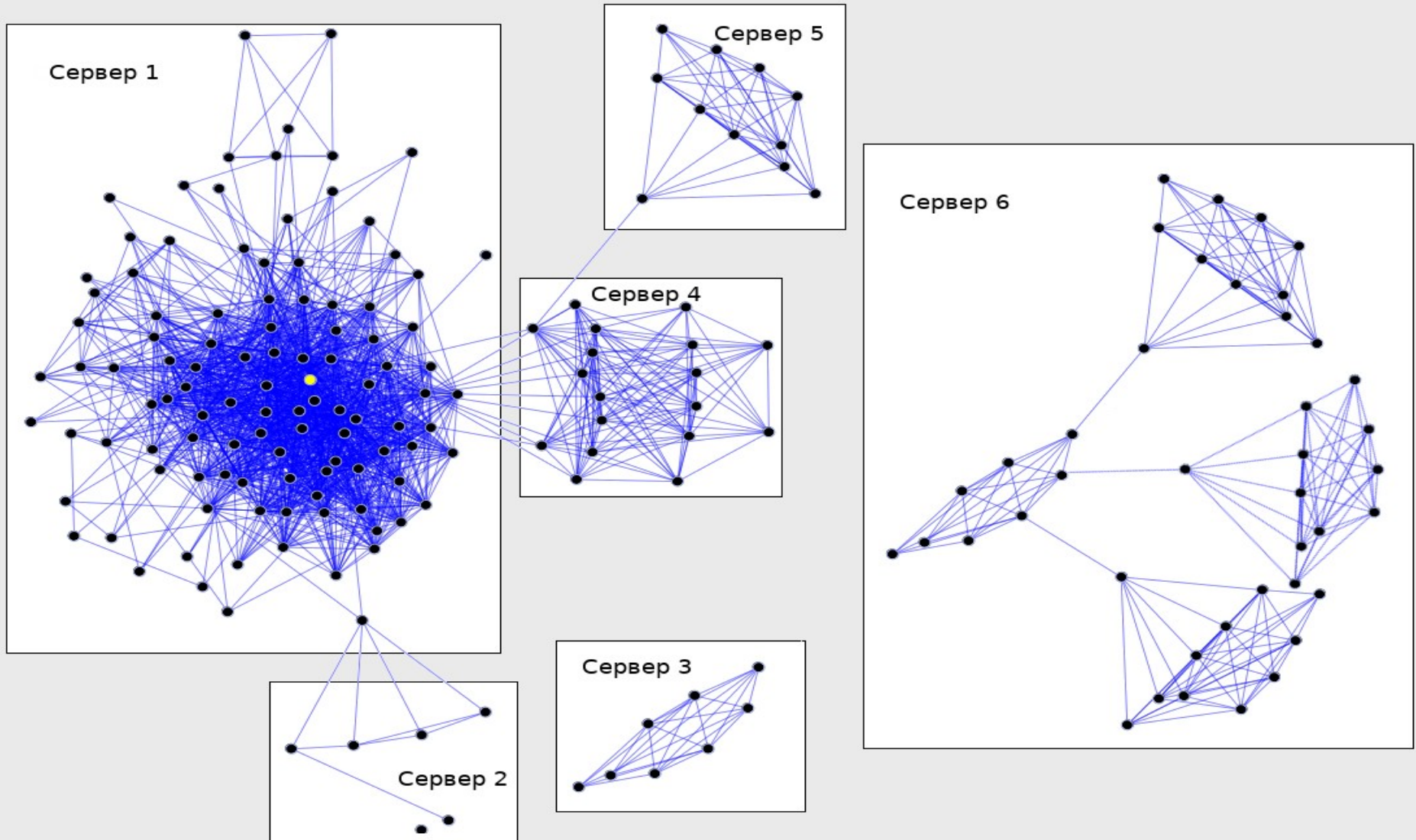
Запись	Ключ (Source)	Имя поля (Property)	Значение поля (Object, Literal)											
USER.ALT_Linux	USER.ALT_Linux													
	<u>USER.ALT_Linux.source=вики</u>	<i>wikiPageWikiLink</i>	[ <a href="#">USER.Дистрибутивы ALT Linux</a> <a href="#">USER.ALT Linux Team</a> <a href="#">USER.ALT Linux (компания)</a> ]											
		<i>wikiPageDisambiguates</i>	[ <a href="#">USER.Дистрибутивы ALT Linux</a> <a href="#">USER.ALT Linux Team</a> <a href="#">USER.Компания ALT Linux</a> <a href="#">USER.ALT Linux (компания)</a> ]											
	<u>USER.ALT_Linux.source=serelex</u>	<i>relation</i>	<table border="1"> <tr> <td><i>value</i></td> <td>0.17194966156248</td> <td><i>word</i></td> <td><a href="#">USER.operating_system</a></td> </tr> <tr> <td><i>value</i></td> <td>0.052439410468004</td> <td><i>word</i></td> <td><a href="#">USER.unix</a></td> </tr> <tr> <td><i>value</i></td> <td>0.012317509707843</td> <td><i>word</i></td> <td><a href="#">USER.window</a></td> </tr> </table>	<i>value</i>	0.17194966156248	<i>word</i>	<a href="#">USER.operating_system</a>	<i>value</i>	0.052439410468004	<i>word</i>	<a href="#">USER.unix</a>	<i>value</i>	0.012317509707843	<i>word</i>
<i>value</i>	0.17194966156248	<i>word</i>	<a href="#">USER.operating_system</a>											
<i>value</i>	0.052439410468004	<i>word</i>	<a href="#">USER.unix</a>											
<i>value</i>	0.012317509707843	<i>word</i>	<a href="#">USER.window</a>											
USER.ALT_Linux_Team	USER.ALT_Linux_Team													
	<u>USER.ALT_Linux_Team.source=вики</u>	<i>subject</i>	[ <a href="#">USER.Категория:ALT Linux</a> ]											
		<i>wikiPageWikiLink</i>	[ <a href="#">USER.Файл:Alt_linux_team_logo.png</a> <a href="#">USER.Рекурсивный акроним</a> <a href="#">USER.Свободное программное обеспечение</a> <a href="#">USER.Сизиф (репозиторий)</a> <a href="#">USER.Дистрибутивы ALT Linux</a> <a href="#">USER.Россия</a> <a href="#">USER.Белоруссия</a> <a href="#">USER.Украина</a> <a href="#">USER.Казахстан</a> <a href="#">USER.Румыния</a> <a href="#">USER.Бразилия</a> <a href="#">USER.ALT Linux (компания)</a> <a href="#">USER.Копилефт</a> <a href="#">USER.Linux</a> <a href="#">USER.Свободный офис</a> <a href="#">USER.Москва</a> <a href="#">USER.2001 год</a> <a href="#">USER.Категория:ALT Linux</a> ]											

# Распределенное хранение/вычисление

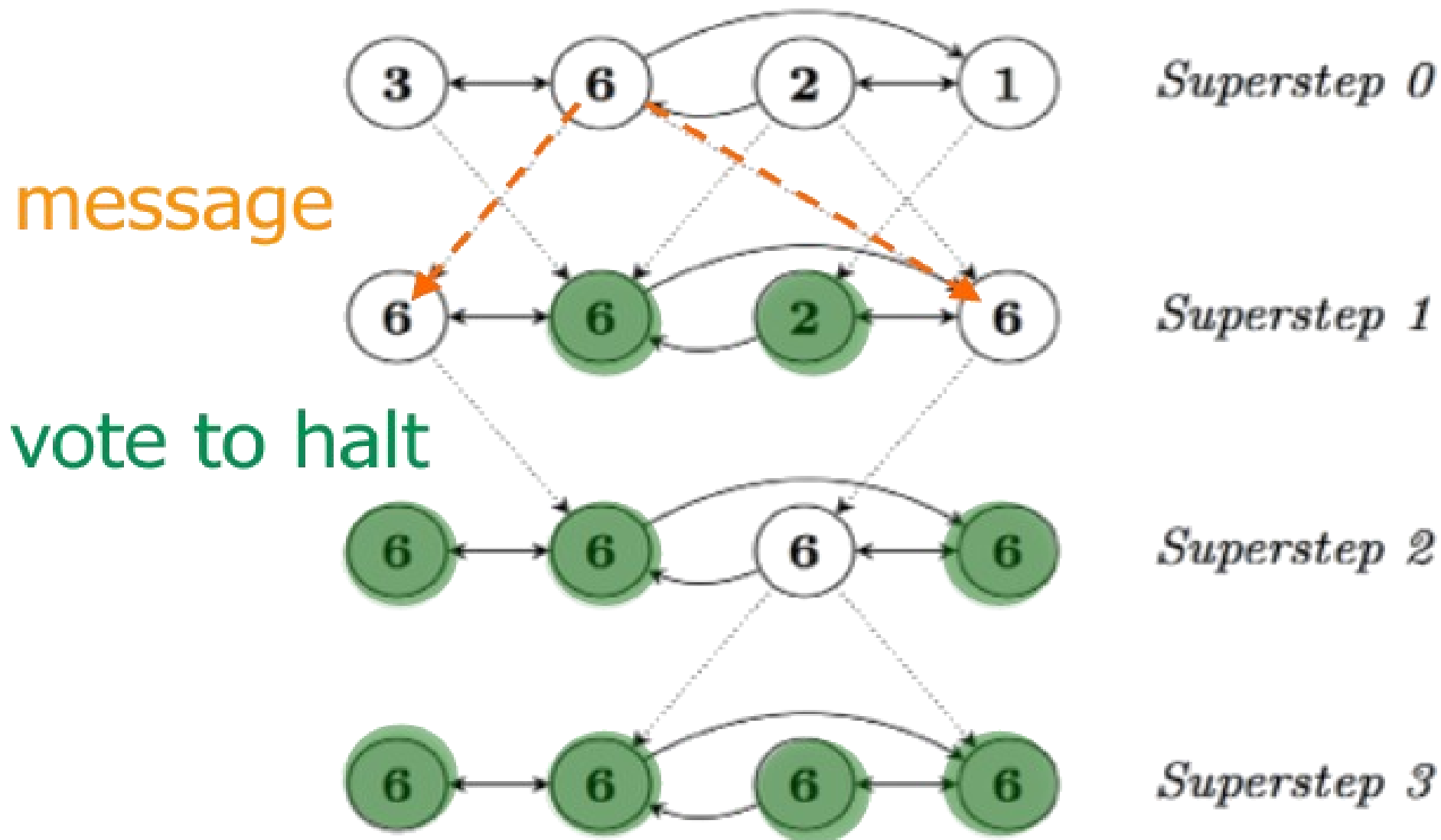




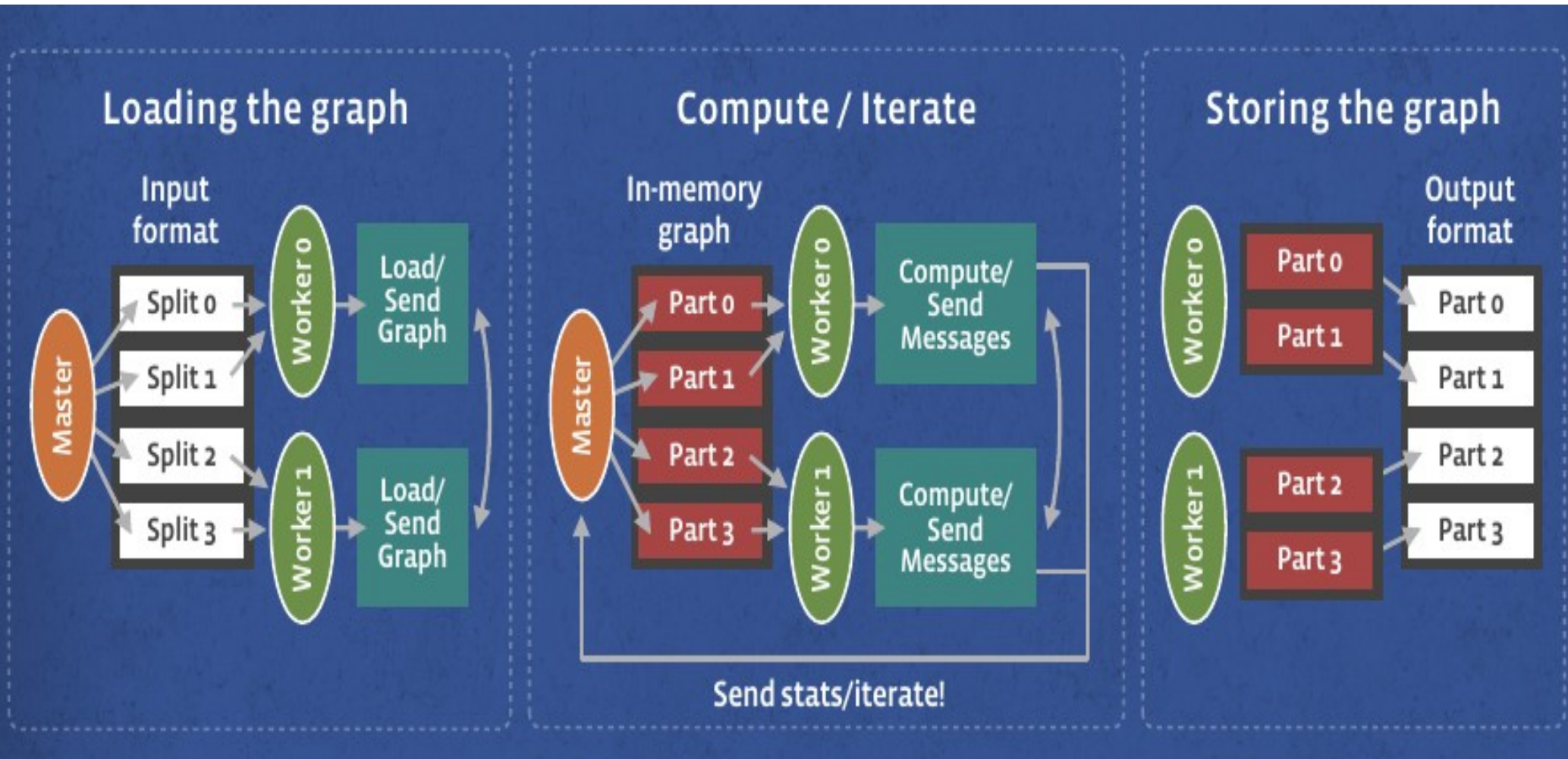
# Apache Giraph (Google Pregel)



# Максимальное значение вершины

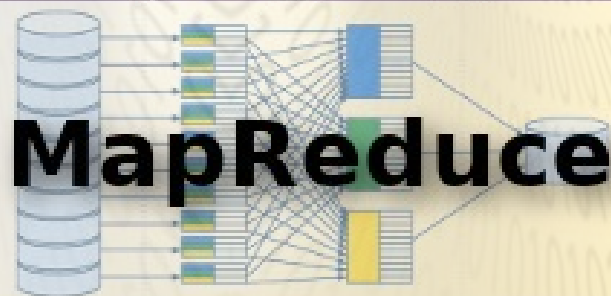


# Поток обработки графов в Giraph



# Стек технологий C2R

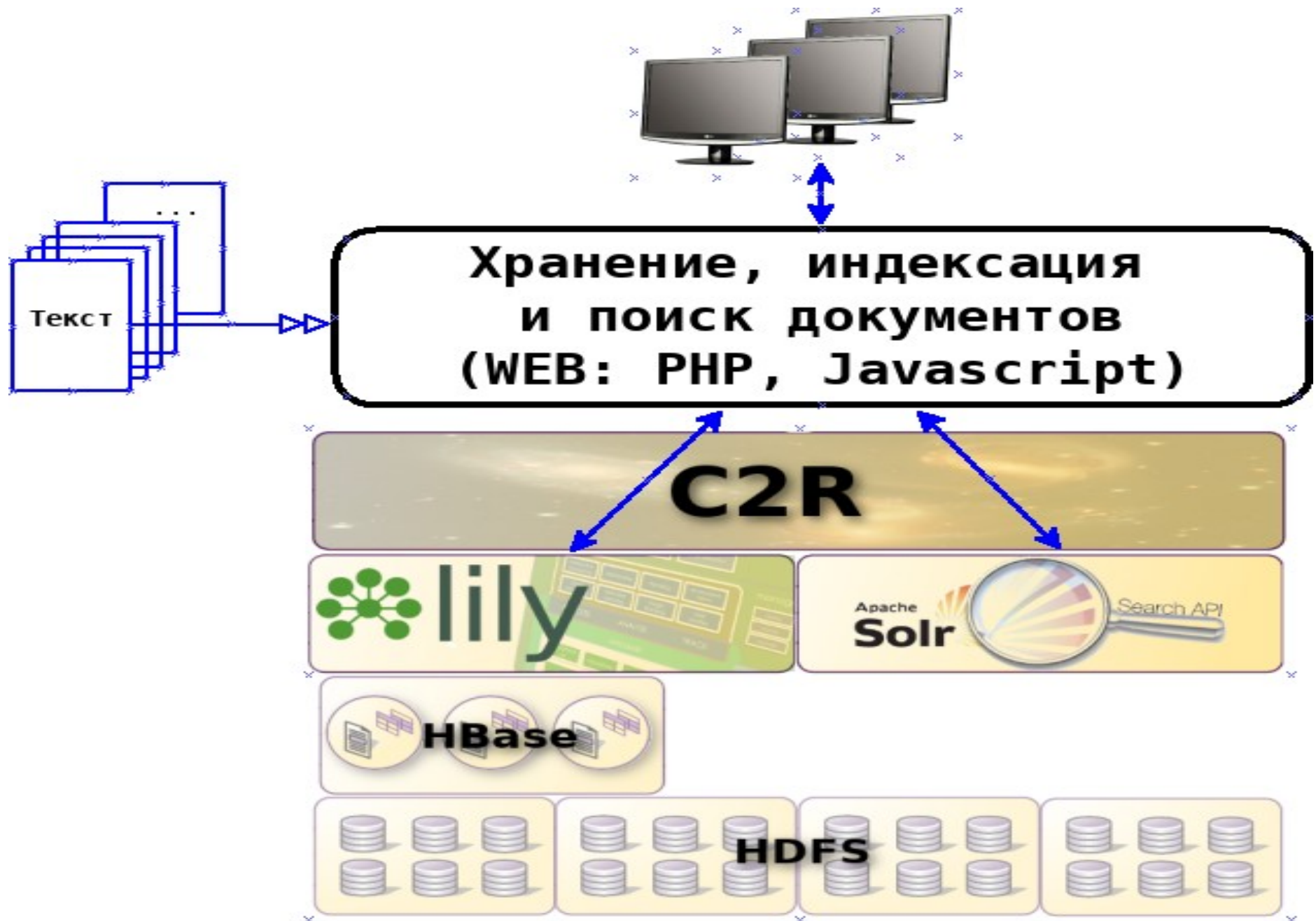
## C2R



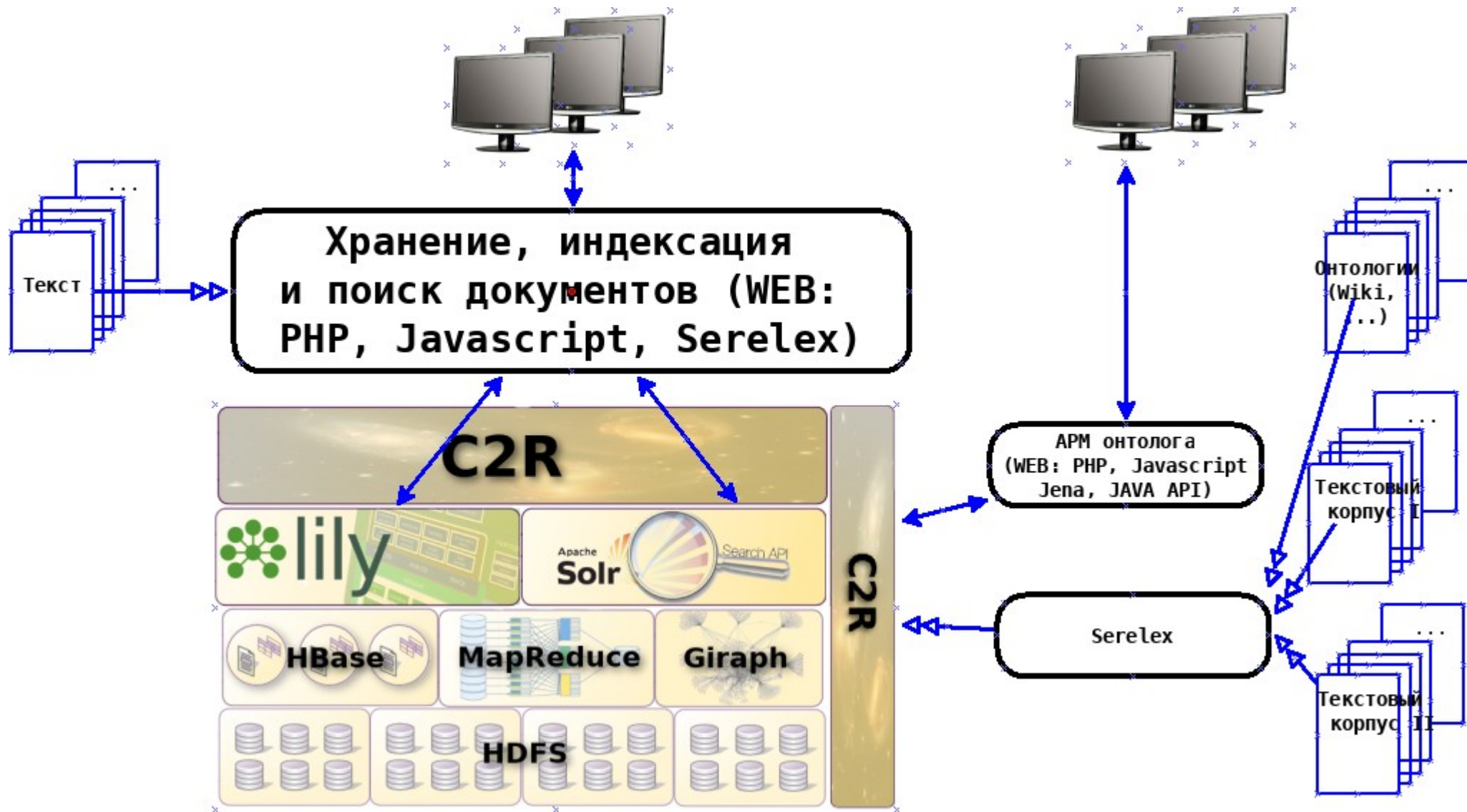
**HDFS**



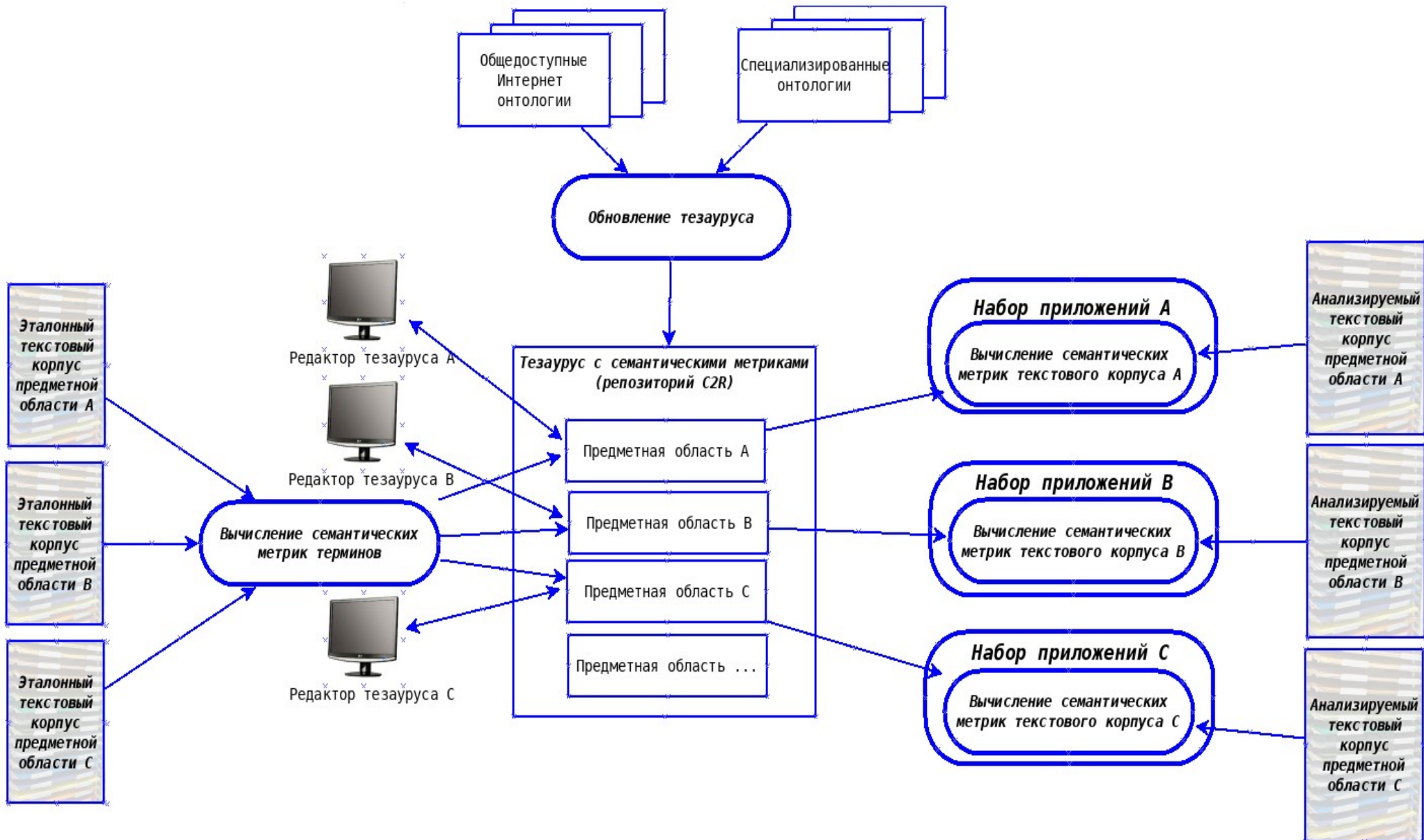
# Прототип системы



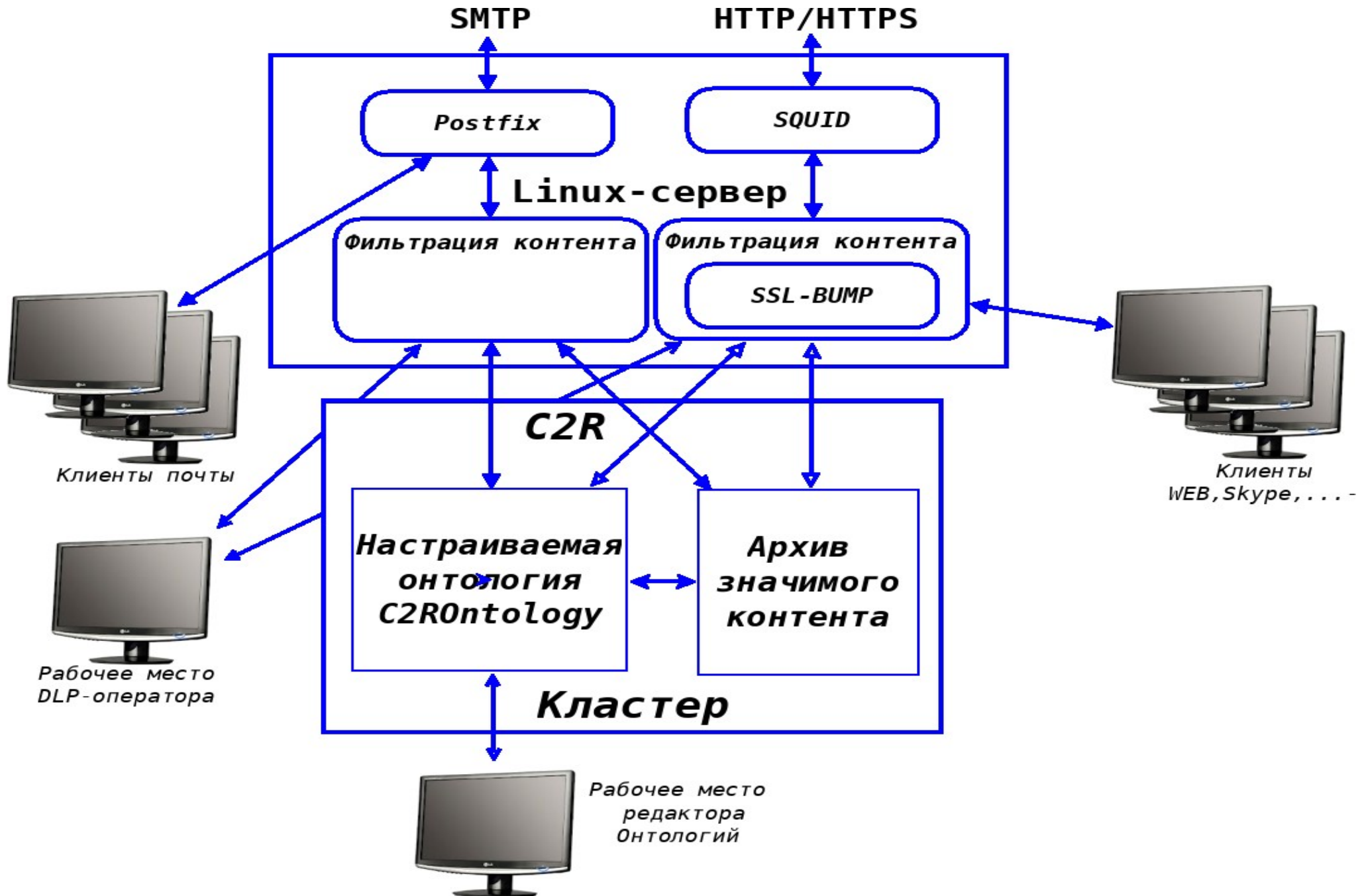
# Полнофункциональное решение



# Информационные потоки C2R Serelex



# DLP-системы





# Спасибо за внимание

Алексей Костарев директор ООО НЕВОД [kaf@nevod.ru](mailto:kaf@nevod.ru)  
Системный аналитик 1-й категории ЛИ ГК ИВС [kaf@ics.perm.ru](mailto:kaf@ics.perm.ru)