



CLAIM
научно-образовательный кластер



Information Technologies of Cognitive Thesauri Design

Andrey Philippovich

04 February 2011



Andrey Philippovich – Work

- ▶ **Ph.D., Prof. of Bauman Moscow State Technical University**
 - Courses: Artificial Intelligence, Computer linguistics and Semiotics, Architecture of Information systems, Design IT-Curriculums
- ▶ **Chief of Science Education Cluster CLAIM** (Computer Linguistics, Artificial Intelligence, Multimedia and more)
- ▶ **Member of Leading Scientific School of Russia “Russian Language Person”** (Head of School – Jury Karaulov)

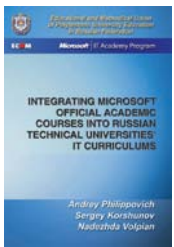
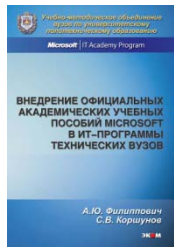
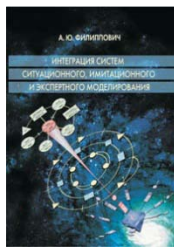
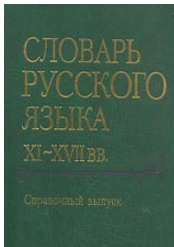


- ▶ Chief of Laboratory of Technical Education in Russia, BMSTU
- ▶ Chief Executive Deputy of Multivendor and Academic ICT Consortium
- ▶ Editor of the Rubric «ICT in Education», Magazine «Quality of Education»
- ▶ IT-Consultant (20+ projects)



Andrey Philippovich– Books

- ▶ Design IT curriculums for HE and VET (2010)
- ▶ Bulletin of MAC ICT (2010)
- ▶ Integrating Microsoft official academic courses into Russian technical universities' IT curriculums (2008)
- ▶ Credit system in Education: automation aspect (2005)
- ▶ Integration of the Situation, Simulation, and Expert Modelling Systems (2003)
- ▶ Computer index of sources of The Hand-written Ancient Cardfile and The Dictionary of Russian Language of the XI–XVII centuries (2002)
- ▶ Information volume of the Dictionary of Russian Language of the XI–XVII centuries. 3 part. Backward wordlist (2001)



Andrey Philippovich– Projects

- ▶ **Information system of associative experiments**

Design methods and tools for automated building associative thesauri, modeling and analysis of language consciousness.



- ▶ **Computer Semiography**

Decoding and systemize russian znamenny (semiographic) chants (XI–XVII centuries). Building musical and idiophone thesauri, ontologies.

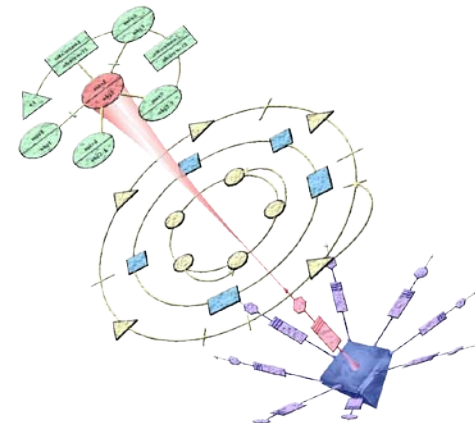


- ▶ **Computer Historical Lexicography**

- The Hand-written Ancient Cardfile of Russian Language of the XI–XVII centuries
- The Dictionary of Russian Language of the XI–XVII centuries
- The Dictionary of the Russian Academy (1789–94)

- ▶ **SIE–modelling Theory**

Design methodology of integration and convergence different decision support systems (Situation, Simulation, Expert, Neurogenetic, Fuzzy, Semiotic and other).



Andrey Philippovich – Projects

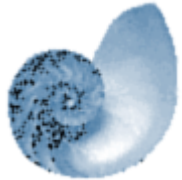
- ▶ **Gesture-mimic interface for interaction with computer**
 - Emotion Recognition based on facial gesture
 - Dictionary of gestures in the field of ICT (audiology)
 - Gesture language recognition

- ▶ **New methods of information visualization (Visual semiotics)**

Implementation and testing of new visualization concepts of familiar information to improve the visibility, convenience and efficiency of its perception and processing.

 - Domain-specific 3D social network
 - Concept of visualization based on the theory of semiotic design of IT
 - Visualization of web-ontologies

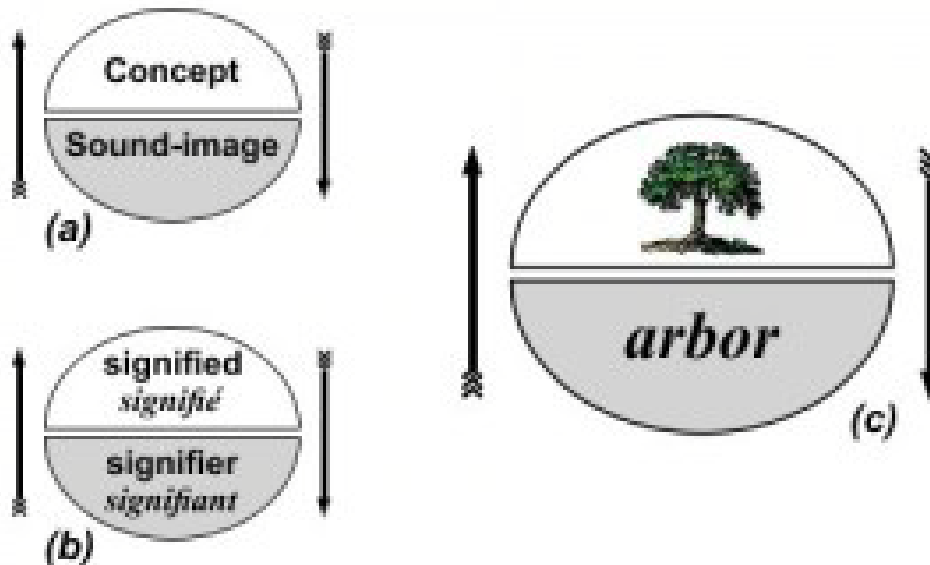
- ▶ **Semantic Web & Text Mining**
 - Ontology design
 - Text and Data clustering
 - Text's tonality analysis



Content

1. **Theoretical background**
 - Semiotics – main ideas
 - Thesaurus and Ontology
 - Cognitive thesauri
2. **R&D Projects**
 - Associative–verbal thesaurus
 - Linguacultural thesaurus
 - Dictionary of metaphors
 - Musical and idiophone «Theons»

Semiotics – main ideas. Sign



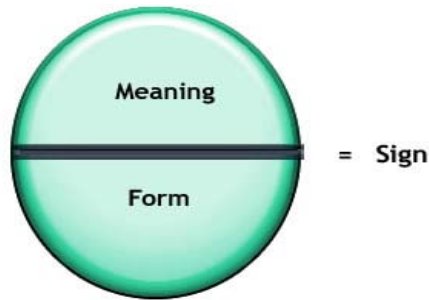
- ▶ The sign is described as a "double entity", made up of the **signifier**, or sound image, and the **signified**, or concept.
- ▶ The basic principle of the **arbitrariness of the sign** in the extract is: there is no natural reason why a particular sign should be attached to a particular concept.

- ▶ The **relationship** between signifier and signified is, however, **not quite that simple**.



Ferdinand de Saussure (1857–1913) was a Swiss linguist and is widely considered to be one of the fathers of 20th-century linguistics and of semiotics

Syntagmatic vs. Paradigmatic

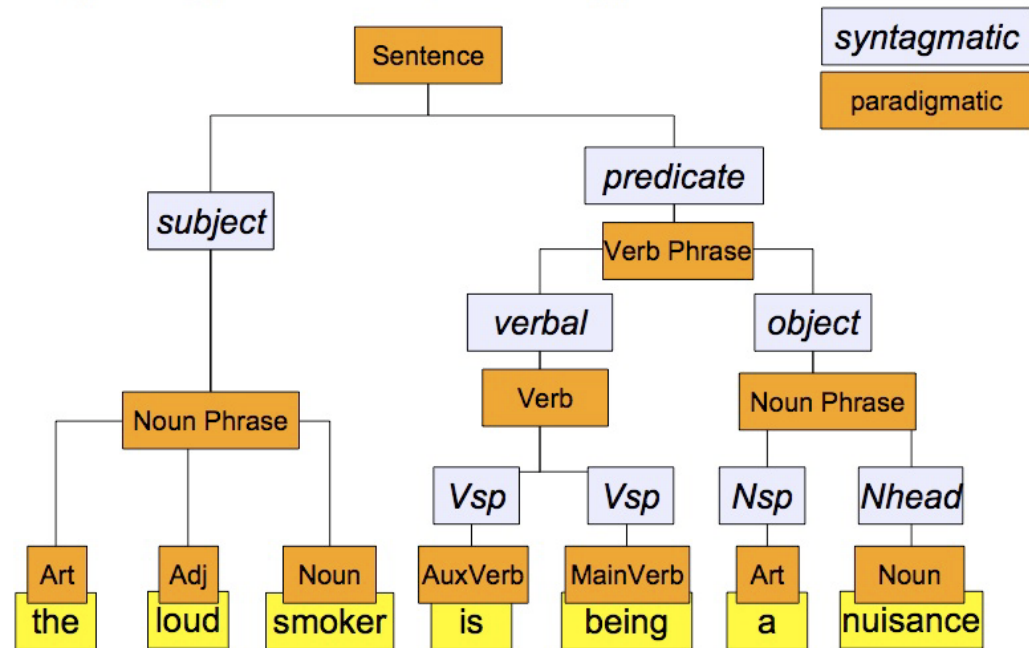


Signified (meaning)

Plane of Content

Plane of Form

Signifier (form)

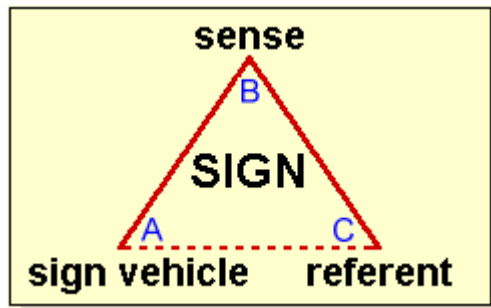
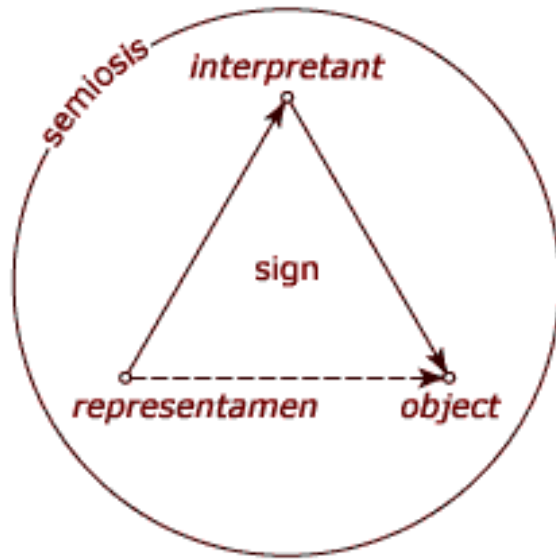


Knowledge element	←←	Paradigmatic relations	→→	Knowledge element
Sign	←←	Syntagmatic relations	→→	Sign

Saussure asserted that there are only two types of relations: **syntagmatic** and **paradigmatic**. The latter is **associative**, and clusters signs together in the mind, producing sets.

Sets always involve a similarity, but difference is a prerequisite, otherwise none of the items would be distinguishable from one another: this would result in there being a single item, which could not constitute a set on its own.

Semiosis – cognitive process

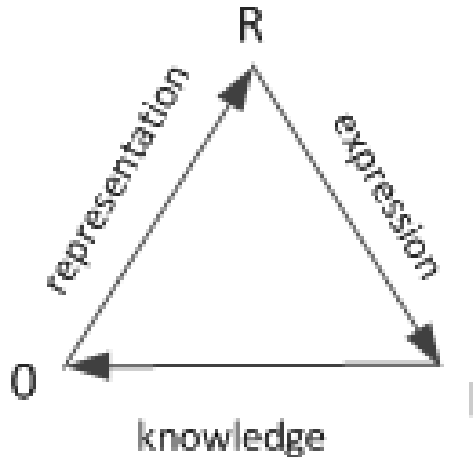


- ▶ **Semiosis** – conceptual (**cognitive**) process, continually unfolding and unending – the **chain of meaning-making** by new signs interpreting a prior sign or set of signs)
- ▶ **Semiosis** as an irreducibly triadic process wherein something, as an **object**, logically determines or influences something as a **sign** to determine or influence something as an interpretation or **interpretant**, itself a sign, thus leading to further interpretants.
- ▶ **Sign vehicle**: the form of the sign
- ▶ **Sense**: the sense made of the sign
- ▶ **Referent**: what the sign 'stands for'



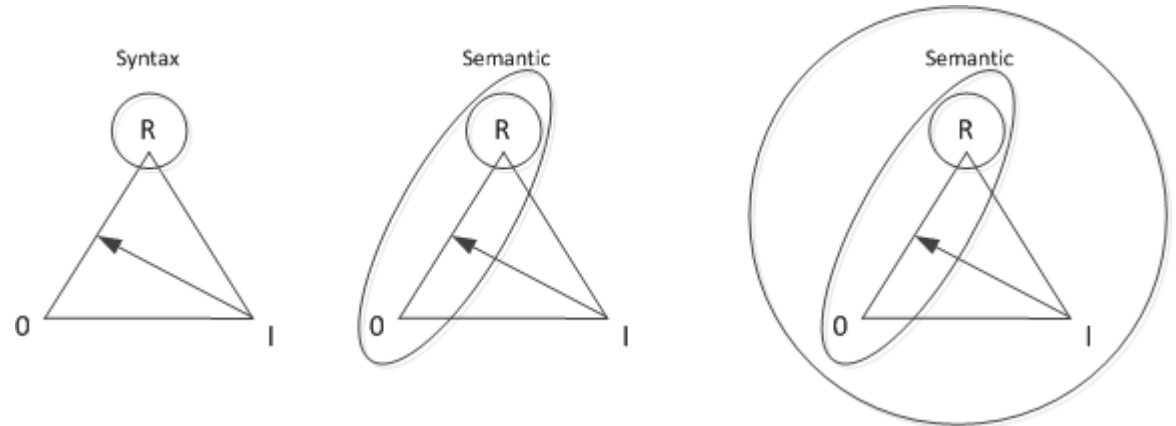
Charles Sanders Peirce (1839 – 1914) was an American philosopher, logician, mathematician, and scientist. He is one of the fathers of semiotics.

Syntax, Semantic and Pragmatic



Sign functions:

- ▶ Representation
- ▶ Expression
- ▶ Knowledge



Three directions of semiotics

- ▶ **Syntax:** the relation between signs, how signs are constituted
- ▶ **Semantic:** the relation between sign and object, what the signs are conveying
- ▶ **Pragmatic:** the relation between signs and the user, what for signs are used

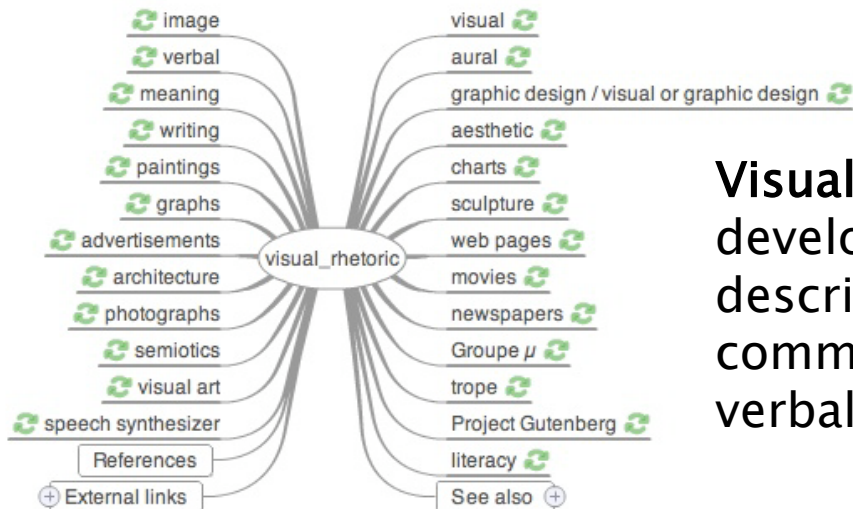


Charles W. Morris (1901 – 1979)
was an American semiotician
and philosopher

Visual semiotics



Visual semiotics – studies of meaning evolve from semiotics, a philosophical approach that seeks to interpret messages in terms of their signs and patterns of symbolism

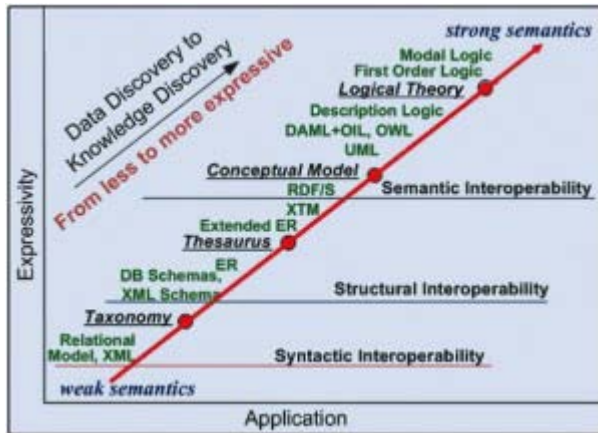


Visual rhetoric is the fairly recent development of a theoretical framework describing how visual images communicate, as opposed to aural or verbal messages.



Groupe μ (founded 1967) is the collective pseudonym under which a group of Belgian 20th-century semioticians wrote a series of books, presenting an exposition of modern semiotics

Thesaurus vs. Ontology



- ▶ These concepts have a lot of definitions, which depends on current subject area, applications in the field of:

- Linguistics
- Information retrieval
- Artificial Intelligence
- Computer semiotics

Artificial Intelligence

- Thesaurus and ontology are some kind of formal models for knowledge representation and management

Information retrieval

- Thesaurus and ontology are some kind of metadata (ISO2788, ISO5964 – thesauri; OWL – ontology)

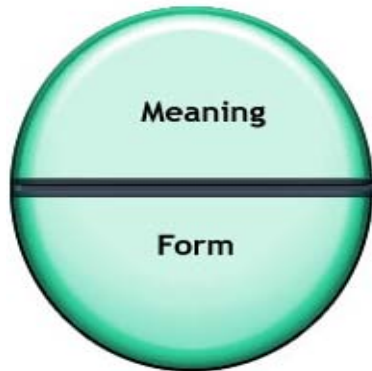
Linguistics

- Thesaurus is most complex lexicographical object. It is something more than dictionary

Computer semiotics

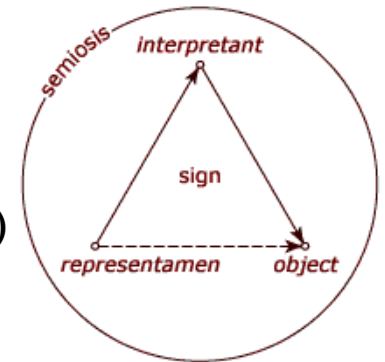
- Thesaurus & Ontology is centaur with name **“THEON”**

Lexical ontology = Theon



Person consciousness

- ▶ language consciousness (Linguistics)
- ▶ Logic consciousness (Pragmatics, AI)
- ▶ Expression consciousness (Psychology)
- ▶ etc.



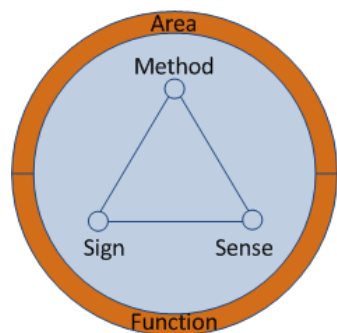
$$\text{LCU} = \text{WKU} + \text{LU}$$

- ▶ LCU – language consciousness unit
- ▶ WKU – world knowledge unit
- ▶ LU – language unit

$$\text{Theon} = \{\text{LCU}\}$$

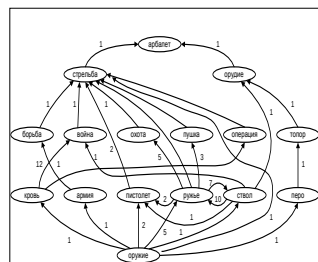
- ▶ **Thesaurus** = {LU}
lexicographical (formal) model of language
- ▶ **Ontology** = {WKU}
Formal model of real world

Cognitive thesauri

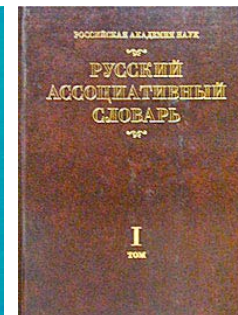
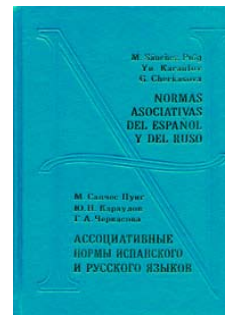
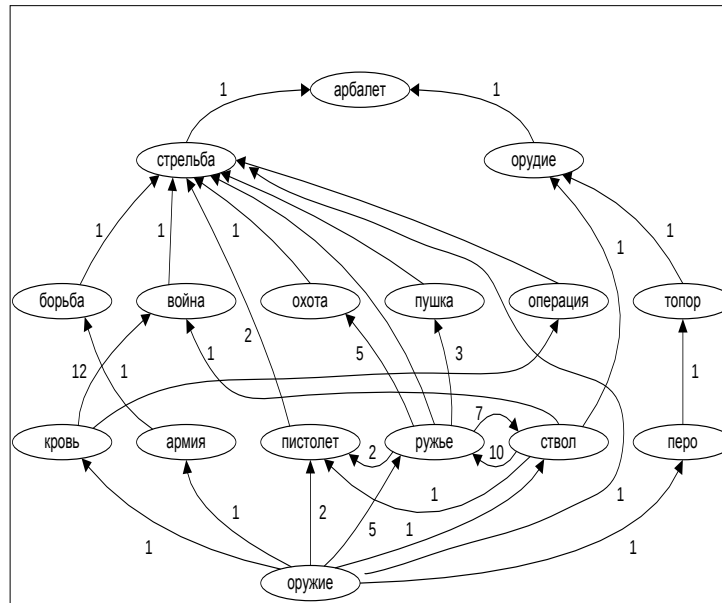


Thesaurus, Ontology ~ ok. What about epistemology?

- ▶ We need new linguistic (and lexicographical) objects, models, grammars, etc.
- ▶ **Cognitive science**, cognitive psychology, cognitive linguistics, cognitive semiotics study how our consciousness works
- ▶ **Cognizer** – cognitive vehicle
- ▶ **Cognitive thesaurus** – representation of Cognizer's activity, structure and dynamics of cognitive processes.
 - ▶ Associative–verbal thesaurus
 - ▶ Linguacultural thesaurus
 - ▶ Dictionary of metaphors
 - ▶ Musical and idiophone «Theons»
 - ▶ Gesture–mimic thesaurus



ASSOCIATIVE-VERBAL THESAURI

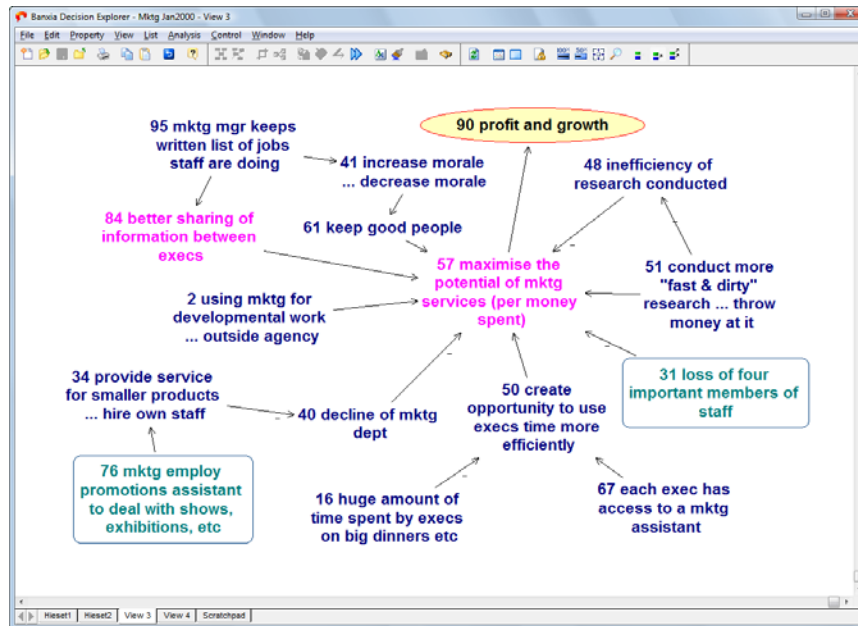


Mental associations

- ▶ **Mental association (Association of Ideas)** – is a term used to refer to explanations about the conditions under which representations arise in consciousness.
- ▶ One idea was thought to follow another in consciousness if it were associated by some principle. **The three most commonly asserted principles of association** were:
 - association by contiguity
 - association by contrast
 - association by similarity
- ▶ The principles of **associationism** were fertile for the progress of psychological investigation – new methods of studying:
 - memory (mechanical—H. Ebbinghaus, Germany; and figurative— F. Galton, England)
 - emotions (C. Darwin, England)
 - motivation (S. Freud, Austria; K. Jung, Switzerland)

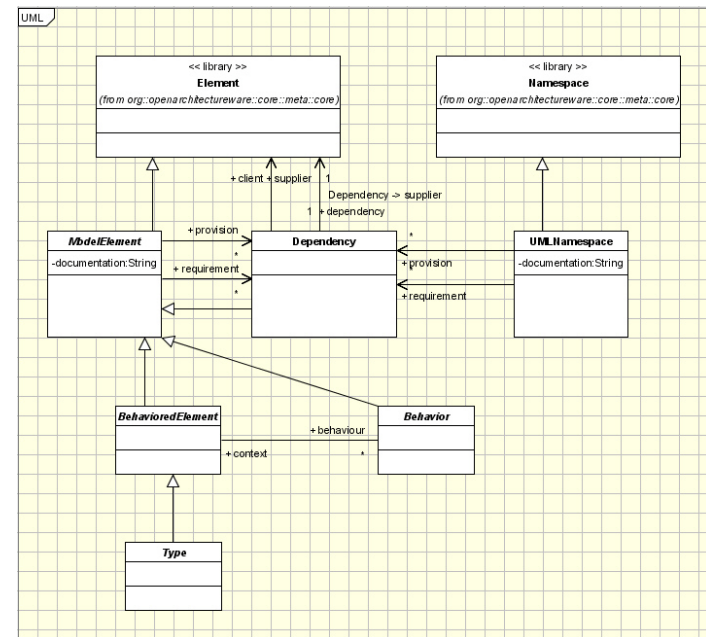
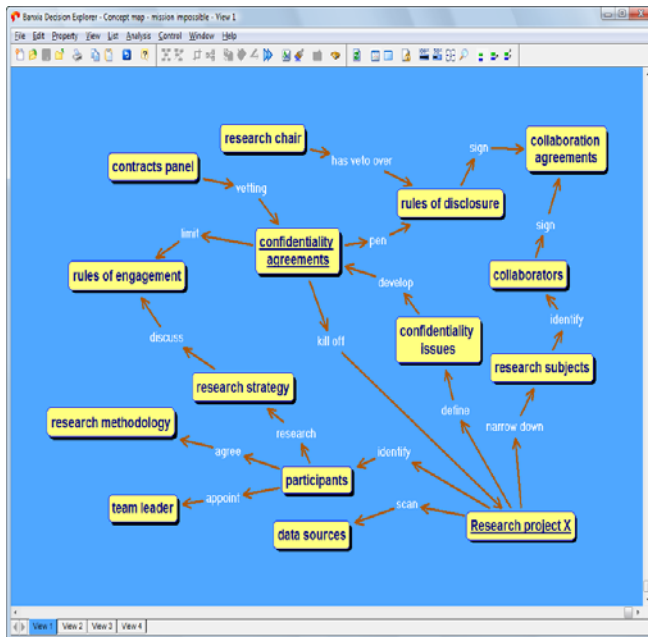
Associations and mindmapping

- ▶ **Association** – The formation of a bond or connection **between ideas** or stimulus and response [*Banerjee J. C. Encyclopaedic Dictionary of Psychological Terms, 1994*].
- ▶ **Mind mapping** (Buzan 1993)
- ▶ **Cognitive mapping** (Eden 1988, 1998, Ackermann et al. 1992)



Associations and Semantic Nets

- ▶ **Association** – undefined or unknown semantic relationship [*Quillian, semantic networks*]
- ▶ In **UML models**, an association is a relationship between two classifiers that describes the reasons for the relationship and the rules that govern the relationship.



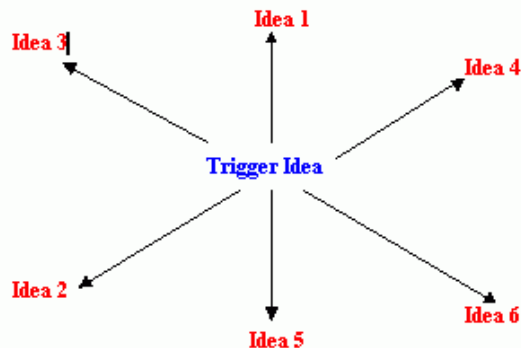
Associative experiment (test)

Le but de cette "expérience" est de récolter des associations d'idées.
En face de chaque mot présenté, vous avez toute liberté d'inscrire ce à quoi il vous fait penser, ce qui vous vient à l'esprit.

âge : 21 sexe (M/F) : M langue maternelle : Français année (1A/2A/3A) : 2A 2 - 0070
heure de début de remplissage : 10h14 date : 6/09/99

14 répertoire...	fichier	41 processeur central...	intel	97 reprogrammable...	oui
3 antivirus...	notion	32 italique...	gras	14 disquette...	petit
57 scanner...	photo	109 archiver...	mémoriser	48 utilisateur...	client
62 programmation...	truc	98 appliqué...	parasol	93 magnétique...	puce
47 recherche...	trouver	40 pirate informatique...	hacker	50 ciel...	mot de passe
49 décryptage...	secret	38 moniteur...	écran	6 octet...	case
12 données...	base	31 copie...	collage	103 distant...	connexion
81 interruption...	plantage	101 téléphonique...	modem	17 lecteur...	disk
77 modification...	paragraphe	28 clavier...	azerty	85 graphique...	carte
22 image...	fig	102 test...	performance	80 alimentation...	secteur
48 restauration...	corbeille	40 souris...	mouse	34 listing...	pas beau
94 formaté...	Netlog	39 multimédia...	cdrom	83 format...	point (.)
4 archive...	mémoriser	44 carte...	Micro	44 périphérique...	imprimante
13 joystick...	jeu	37 modem...	Oltec	84 formatage...	nettoyage
104 physique...	composants	45 pixel...	fibre	121 régler...	paramètres
11 driver...	delix	20 disque dur...	giga	71 plantage...	bug
53 procédure...	programme	99 réseau...	hub	115 redémarrer...	Alt F4
114 visualiser...	potentiomètre	7 banque de données...	archive	27 cartouche...	couleur
50 imprimante...	couleur	2 algorithme...	boucle	106 numérique...	acquisition
49 ordinateur portable...	cher	89 interactif...	game	30 ordinateur...	babare

Expérience linguistique associative - sept.1999 - MGU/ECL - Dr. Pr. Iouri N. Philippovitch/Denis Dheft
heure de fin de remplissage : 10h31



- ▶ **Associative experiment (AE) or association test** – one of the most widespread methods psycholinguistics and used to study the organization of mental life, with special reference to the **cognitive connections** that underlie perception and meaning, memory, language, reasoning, and motivation.

$$\{S\} \rightarrow \{R\}$$

- ▶ In the **free-association test** the subject is told to state the first word (**Response** or **Reaction**) that comes to mind in response to a stated word, concept, or other **Stimulus**.
- ▶ In **controlled association test** a relation may be prescribed between the **Stimulus** and the **Response** (e.g., the subject may be asked to give opposites).

Articles of Associative Dictionary

Interface (S)

- ▶ {R} 7: graphique 5: lien 4: homme-machine 2: écran; imprimante 1: acquisition; arrivée; barrière; candide; communication; connexion; conviviale; de liquides; double; driver; entre; icônes; informatique; limite, séparation; liquide; milieux; MIS; moyen; musicale; ordinateur; parallèle; Pentium III; périphérique; prisme; processeur; raccord; sortie; thermodynamique; traitement de surface; utilisateur; visible; win 95; Windows (o68, d38, v15)

Internet (S)

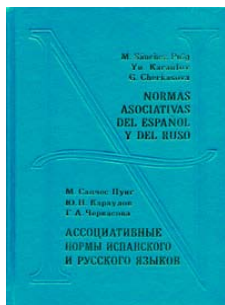
- ▶ {R} 11: web 7: réseau 5: mail 3: e-mail; liberté 2: chat; Netscape 1: AOL; base de données interactive; bof; communication; du cul, du cul, du cul; échange; espace; explorer; fouillis; gratuit; inet; info; intranet; le boom des années 90; lent; loin; MIS; modem; océan; oui; pénible; réseau des réseaux; site; surf; surfons; toile; un www personnel; voyager; www (o68, d36, v6)



**Associations in information technologies:
russian–french experiment (1999–2000)**

http://philippovich.ru/Library/Books/Association_IT/CONTENTS.HTM

Associative dictionaries and thesauri



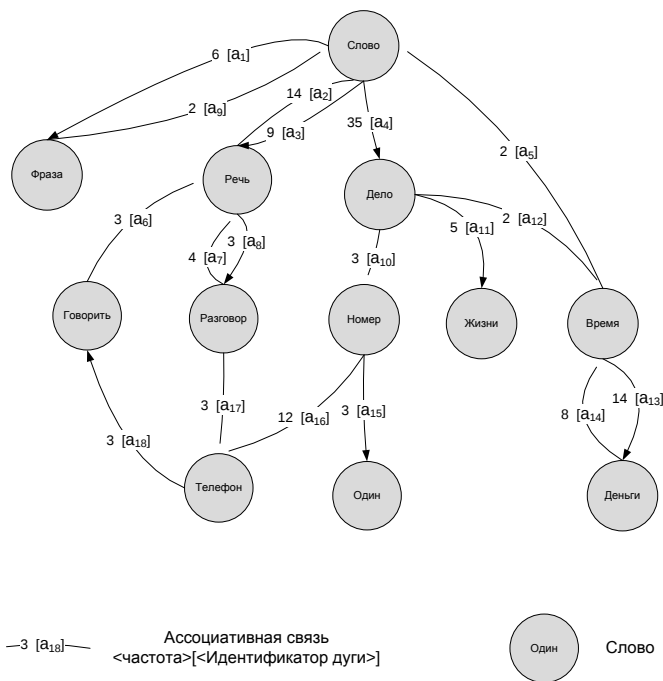
- ▶ The Dictionary of associative norms of Russian of A.A. Leontev (25 thousand records, 1967–1973)
- ▶ The Russian associative dictionary / thesaurus (1,3 million records, 1988–1995)
- ▶ The Slavic associative dictionary (Russian part consist of 66 thousand records, 1998–1999)
- ▶ Associations of information technologies: experiment in Russian and French languages (12,6 thousand records, 1998–2000).



- ▶ Word Association, rhyme and fragment norms (750 thousand records, 1999, The University of South Florida)
- ▶ The Edinburgh Associative Thesaurus (840 thousand records, 1969–1971)



Associative-verbal Model



- **Associative-verbal field (AVF)**
set of all mental associations (reactions) for word

$$AVF(S) = \{R_i\}, R_i \ni S \rightarrow \{R\}$$

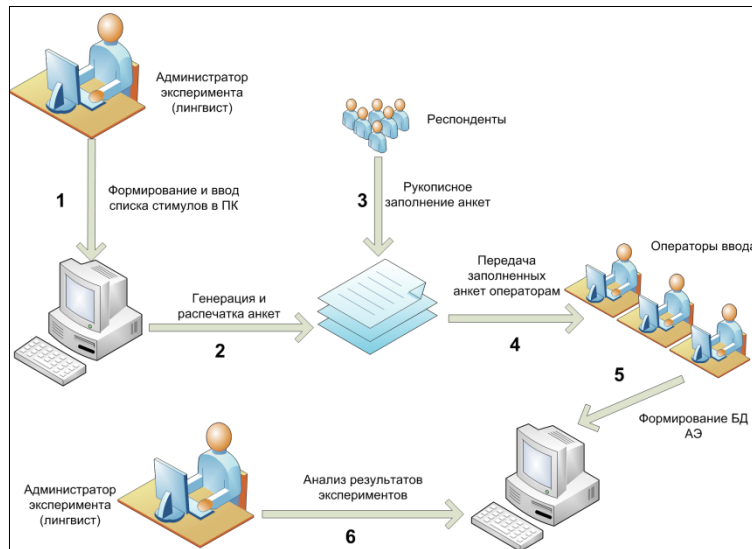
- **Associative-verbal Network (AVN)**
network, where nodes – stimuli and reactions, arcs – associative links.

$$AVN = \langle \{S\} \cup \{R\}, \{\leftarrow, \rightarrow, \leftrightarrow\}, \mathcal{R}(S \rightarrow R) \rangle$$

- **Hypothesis:**
AVM – is cognitive model of language consciousness, main part of semiosis

- **Associative-verbal model (AVM)**
phenomenological model of AVN of persons represented by associative thesaurus or dictionary

Methodology of associative-verbal thesaurus design



1. Analysis prerequisites, conditions and goals of AE
2. Creating list of word-stimulus
3. Generating the set of forms (questionnaires)
4. Questionnaire design
5. Design and filling linguistics database of AE
6. Explore and identification relationships between language constructions
7. Creating thesaurus and linguistics database of AE
8. Estimating thesaurus for necessity of additional questionnaire design (iterational repeating steps 2-7).

The Russian associative experiment



Type:	one free association
Number of stimuli:	6,624
Number of forms:	11,500
Number of stimuli per form:	100
Number of participants:	11,000
Time for filling one form:	7–10 minutes
Number of S→R pairs:	1,032,522
Number of different S→R pairs:	462,500
Number of different reactions:	102,926

Step	Stimuli	Stimuli per subject	Forms	
			Printed	Entered
I (+extra)	1277	100	1500	~1300
	760	~400	3500	~3200
II	2685	100	3000	~2800
III	2935	100	3500	~3150

Electronic edition of Russian Associative thesaurus on CD (2002 –2004)



SR Русский Ассоциативный Слова...

человек R-S G A F

Реакции	Частота
невидимка	25
хороший	25
амфибия	22
добрый	21
разумный	18
животное	16
умный	16
зверь	15
обезьяна	13
большой	12
друг	10
дела	9
люди	9
/	8
земля	8
любимый	8
существо	8
гордый	7
я	7

644 286 8 204

- ▶ Desktop application (Delphi + Paradox)
- ▶ Frequency/Alphabet order
- ▶ Forward / Backward mode ($S \rightarrow R$ / $R \rightarrow S$)
- ▶ Gender comparative mode
- ▶ Normalization for comparatives
- ▶ Gender/Age/Profession filters

SR Обратный PAC: Гендерное представление

язык C-Я S-R G A F

Стимулы	Частота	Мужчины	Женщины
иностранн	311	79	232
родной	226	80	146
английский	67	40	27
французский	61	32	29
национальный	45	8	37
украинский	42	14	28
немецкий	41	16	24

1609/607/1000 238/146/167 138/85/107

Automated system of scientific researches AE



The project of **Russian Foundation for Humanities** (2006–8)

"The Automated system of scientific researches of dynamics of associative–verbal model of language consciousness of Russians as indicator of an image of Russia in the newest history and the present"



- ▶ **Theoretical basis** – hypothesis that results of AE allow :
 - to estimate features of perception of the person
 - to study its language consciousness by means of AVM



Integrated linguistic database based on :

- The Russian associative dictionary
- The Dictionary of associative norms of Russian of A.A. Leontev
- The Slavic associative dictionary
- Associations of information technologies: experiment in Russian and French languages

ASSR AE –Main Results



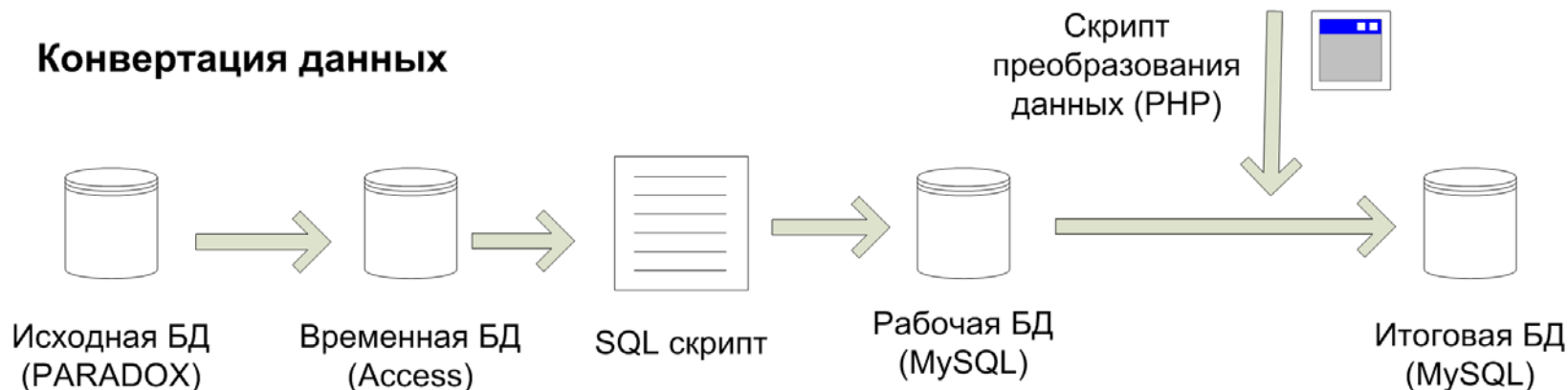
- ▶ **Web-services, program tools and resources:**
 - Hypertext versions of dictionaries
 - Specialized web-version of Russian associative thesaurus
 - System for organizing interactive associative experiment in web
 - Program modules for searching associative chains



- ▶ **New dictionaries & publications**
 - The comparative associative dictionary of Russian
 - Gender Russian associative thesaurus /dictionary
 - 20+ scientific publications
 - 2 master degree projects

Web-version of Russian associative thesaurus

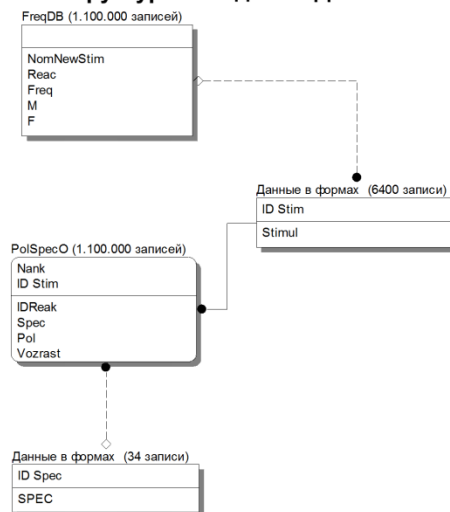
Конвертация данных



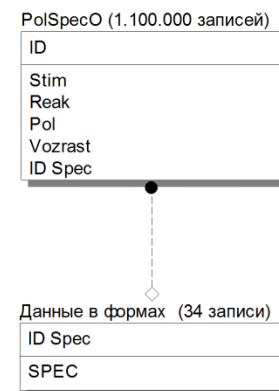
Временные характеристики

ЭТАП	ТИП ЗАПРОСА	КОЛИЧЕСТВО ЗАПРОСОВ	ЗАТРАЧЕННОЕ ВРЕМЯ, ч.
Импорт Paradox-Access	INSERT	3.100.000	12
Access-SQL dump - MySQL	INSERT	3.100.000	2
Слияние таблиц	UPDATE	6.400	18

Структура исходной БД



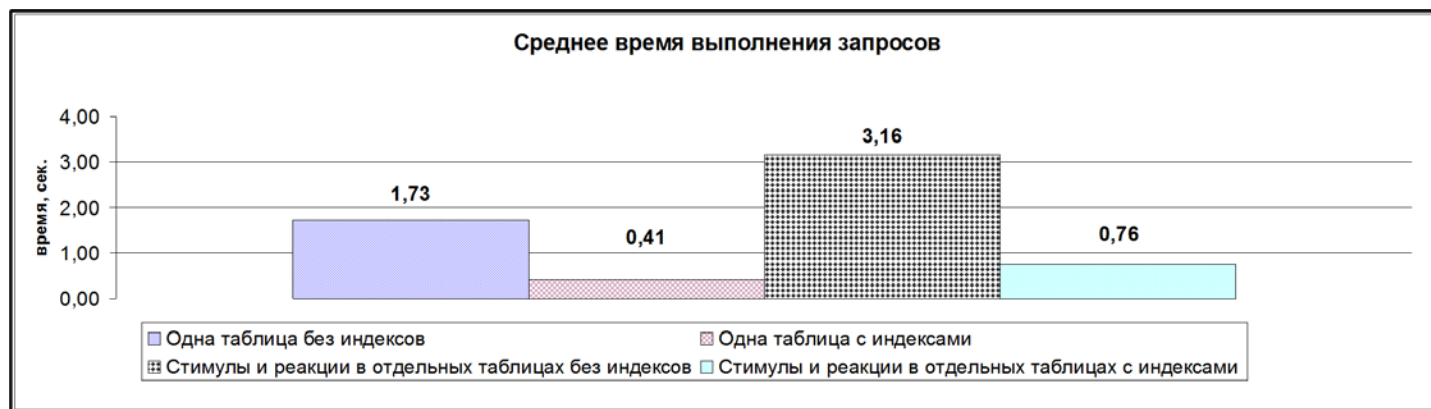
Конечная структура БД



<http://philippovich.ru/Projects/ASIS>

Web-version of RAT – query performance

Тип	Стимул (реакция)	Вид поиска	Кол-во реакций (стимулов)	Время, сек.
Одна таблица без индексов	человек	прямой	644	1,40
	экзамен	прямой+сравнение	507	2,40
	человек	обратный	12654	1,90
	экзамен	обратный+сравнение	681	1,20
			Среднее время	1,73
Одна таблица с индексами	человек	прямой	644	0,04
	экзамен	прямой+сравнение	507	0,12
	человек	обратный	12654	0,54
	экзамен	обратный+сравнение	681	0,93
			Среднее время	0,41
Стимулы и реакции в отдельных таблицах без индексов	человек	прямой	644	2,17
	экзамен	прямой+сравнение	507	3,21
	человек	обратный	12654	3,17
	экзамен	обратный+сравнение	681	4,10
			Среднее время	3,16
Стимулы и реакции в отдельных таблицах с индексами	человек	прямой	644	0,17
	экзамен	прямой+сравнение	507	0,39
	человек	обратный	12654	1,97
	экзамен	обратный+сравнение	681	0,49
			Среднее время	0,76



Web-version of RAT – forms

Русский ассоциативный словарь - Windows Internet Explorer

http://tesaur...

Файл Правка Вид Избранное Сервис Справка

Избранное MAC_ICT_Live CLAIM MAK ИКТ SkyDrive

Веду... Рус... x

Страница

Русский ассоциативный словарь

Запрос: машина

Вид поиска:
☒ Прямой поиск
☐ Обратный поиск

Возраст от 10 до 100

Профессия
 все

Пол ☐ любой ☐ только мужчины ☐ только женщины ☒ сравнение полов

Сортировка ☒ По частоте ☐ По алфавиту

Нормализация ☐ Включить нормализацию по частоте

Поиск

Статистика по запросу:

всего реакций на стимул: 490,
 различных реакций на стимул: 193,
 одиночных реакций на стимул: 148,
 отказов: 3.

Интернет | Защищенный режим: вкл.

Русский ассоциативный словарь - Windows Internet Explorer

http://tesaur...

Файл Правка Вид Избранное Сервис Справка

Избранное MAC_ICT_Live CLAIM MAK ИКТ SkyDrive

Веду... Рус... x

Страница

Статистика по запросу:

всего реакций на стимул: 490,
 различных реакций на стимул: 193,
 одиночных реакций на стимул: 148,
 отказов: 3.

У мужчин:

всего реакций на стимул: 154,
 различных реакций на стимул: 87,
 одиночных реакций на стимул: 69,
 отказов: 2.

У женщин:

всего реакций на стимул: 336,
 различных реакций на стимул: 138,
 одиночных реакций на стимул: 105,
 отказов: 1.

Реакции на стимул "машина":

Реакции	Частота	Мужчины	Женщины
времени	125	32	93
легковая	30	6	24
жигули	23	3	20
Волга	12	2	10
автомобиль	11	6	5

Интернет | Защищенный режим: вкл.

Русский ассоциативный словарь - Windows Internet Explorer

http://tesaur...

Файл Правка Вид Избранное Сервис Справка

Избранное MAC_ICT_Live CLAIM MAK ИКТ SkyDrive

Веду... Рус... x

Страница

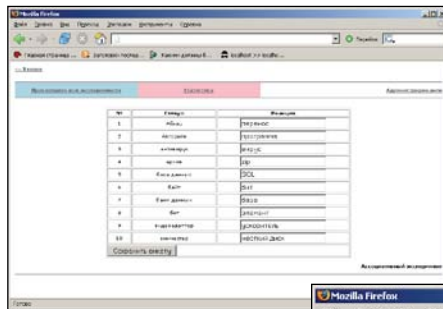
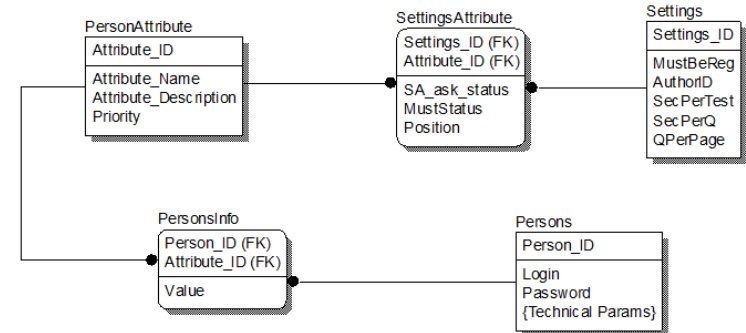
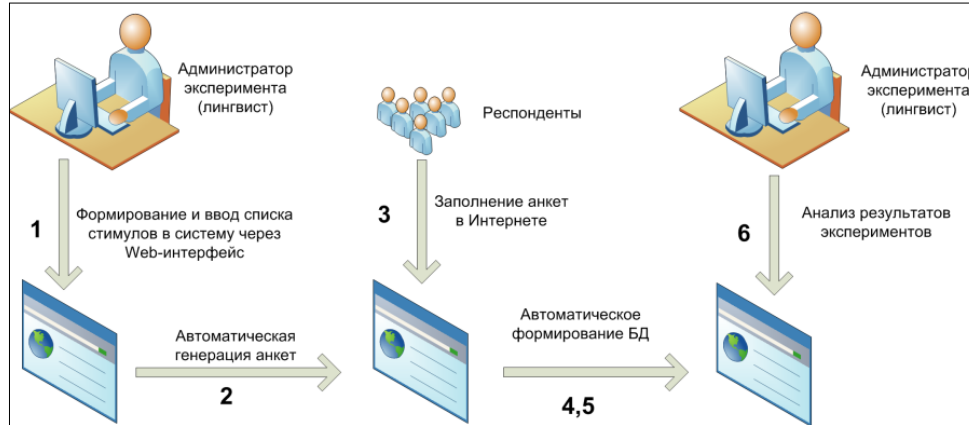
Стимулы, вызывающие реакцию "купить":

Стимулы	Частота	Мужчины	Женщины
дешево	26	16	10
достать	14	5	9
продать	14	6	8
приобрести	11	8	3
стоит	8	3	5
вещь	6	1	5
выгодно	5	3	2
успеть	4	1	3
билет	3	0	3
велосипед	3	1	2
заплатить	3	1	2
магазин	3	1	2
просила	3	0	3
что-либо	3	1	2
возможность	2	1	1
магазину	2	1	1
мебель	2	2	0
мыло	2	0	2
отнять	2	0	2
пальто	2	1	1
попросила	2	0	2
поступить	2	1	1
плываю	2	0	2

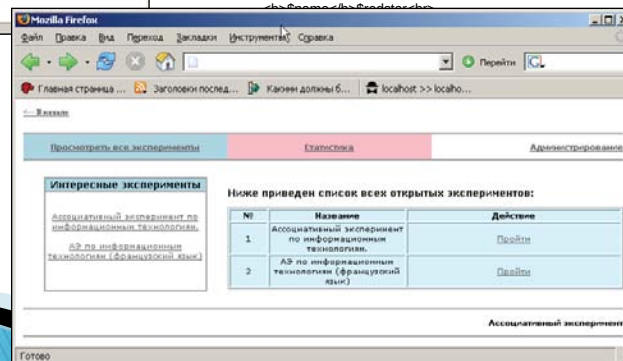
Интернет | Защищенный режим: вкл.

<http://philippovich.ru/Projects/ASIS>

Web-service for interactive AE



```
$query="SELECT *
FROM settingsattribute sa, personattribute pa
WHERE pa.Attribute_ID = sa.Attribute_ID
AND sa.Settings_ID=(SELECT Settings_ID FROM experiment WHERE Experiment_ID=1)
AND sa.Ask_Status=1
ORDER by sa.Position";
$result=mysql_query($query);
$num=mysql_num_rows($result);
for ($i=1;$i<=$num;$i++){
    $row=mysql_fetch_array($result);
    $name=$row['Attribute_Name'];
    $desc=$row['Attribute_Description'];
    $must=$row['MustStatus'];
    $redstar="";
    if ($must==1) $redstar="<font color=red><font>";
    $ID=$row['Attribute_ID'];
    print "<tr><td>";
```



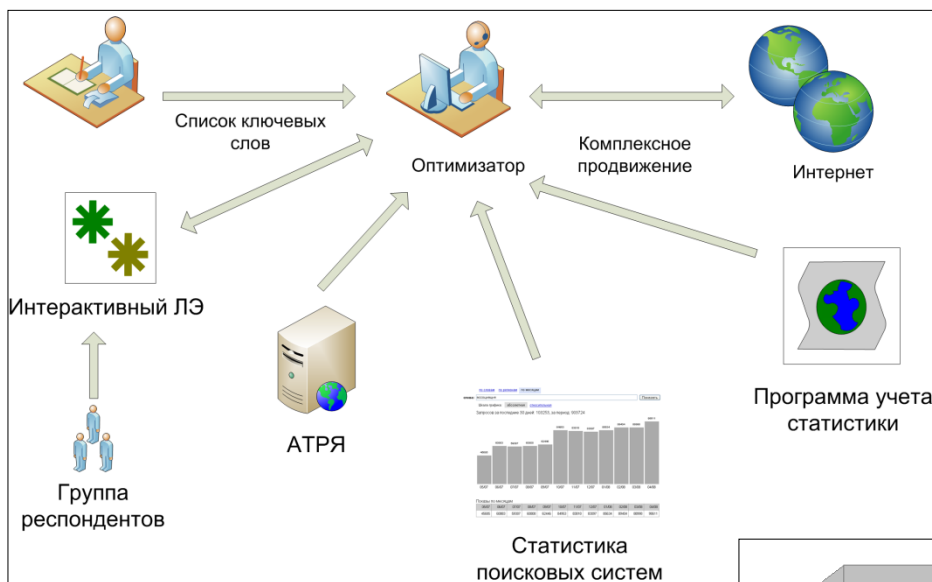
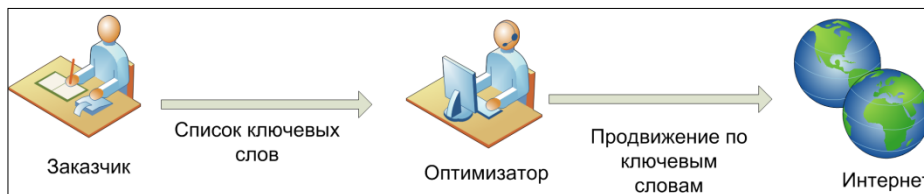
Изменение атрибута:

Описание	Значение
Название Пример: "Страна проживания"	Имя
Описание Описание выводится пользователю рядом с атрибутом Пример: "Введите страну в которой Вы сейчас проживаете"	Введите свое имя, пример
Положение в списке Определяет взаимное расположение атрибутов (только при настройке экспериментов, при прохождении анкетирования расположение индивидуально для каждого эксперимента)	1
Сохранить атрибут	

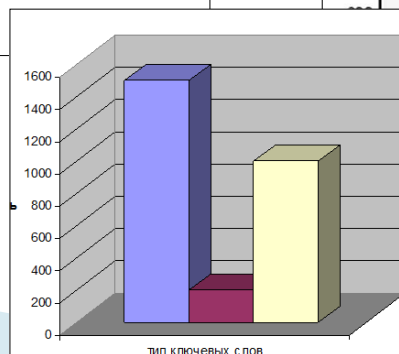
Изменение настроек:

Системные настройки эксперимента:	
Описание	Значение
Для прохождения теста респондент должен быть зарегистрирован в системе	☒
Введите количество вопросов (стимулов) в анкете (штук).	10
Требуемые характеристики респондента:	
Имя Введите свое имя, пример "Иван"	☒ Запрашивать характеристику ☐ Сделать обязательной для заполнения ☐ Положение в списке
Фамилия Введите свою фамилию, пример "Петров"	☒ Запрашивать характеристику ☐ Сделать обязательной для заполнения ☐ Положение в списке
Отчество Введите свое отчество, например "Борисович"	☒ Запрашивать характеристику ☐ Сделать обязательной для заполнения ☐ Положение в списке

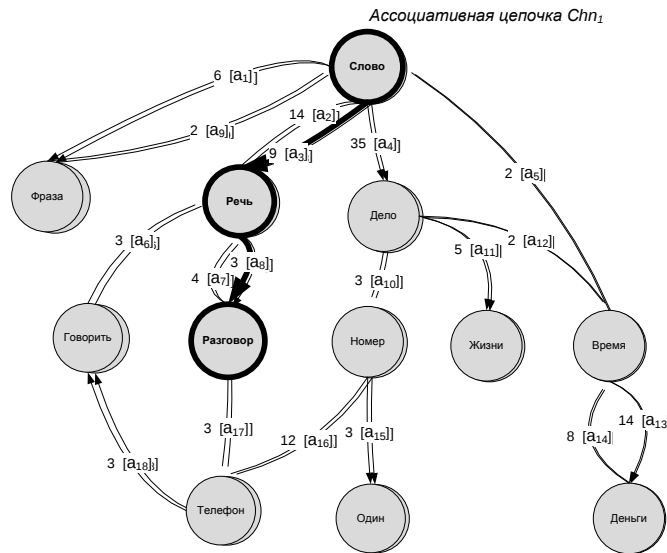
Web-sites optimization based on AE



- ▶ AE makes possible to find non-evident key phrases and collocations
- ▶ It's possible to check sets of stimuli and reactions for keywords
- ▶ Best results – using domain-specific AE, Interactive AE



Associative-verbal chains

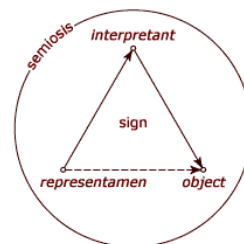


- ▶ Associative-verbal Network (AVN) network, where nodes – stimuli and reactions, arcs – associative links.
 $AVN = \langle \{S\} \cup \{R\}, \{\leftarrow, \rightarrow, \leftrightarrow\}, \mathcal{R}(S \rightarrow R) \rangle$

- ▶ Associative-verbal chain (AVC) ordered set of nodes and arcs based on AVN

$$AVC = S_1 \rightarrow (R_1 = S_2) \rightarrow \dots \rightarrow R_n$$

Ассоциативная связь
 $\leftarrow 3 [a_{18}] \leftarrow$
 $\leftarrow 3 [a_{18}] \leftarrow$

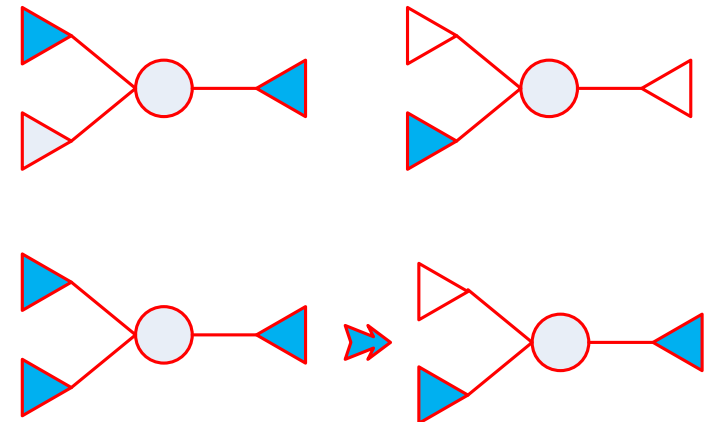
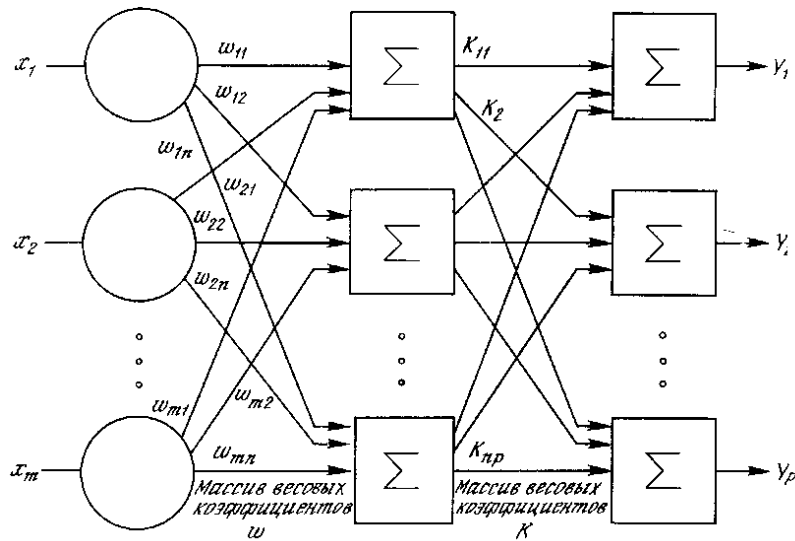
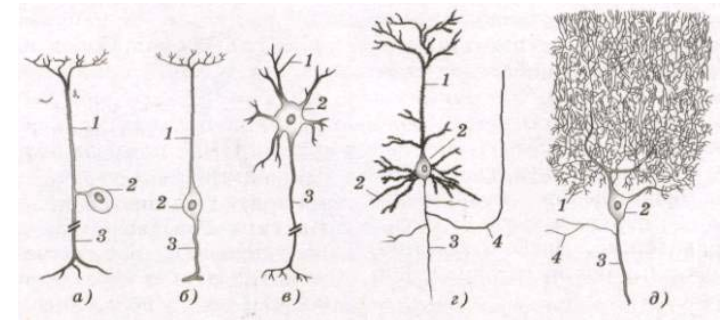
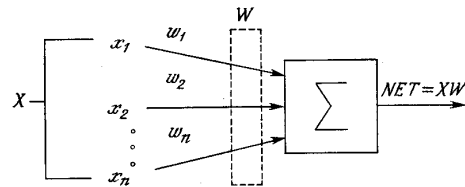
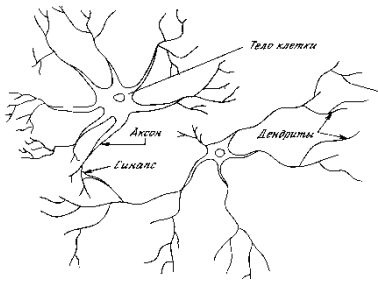


Hypothesis:

- ▶ Producing AVC is a cognitive process, main part of semiosis
- ▶ This process is not simple, because of spreading activation

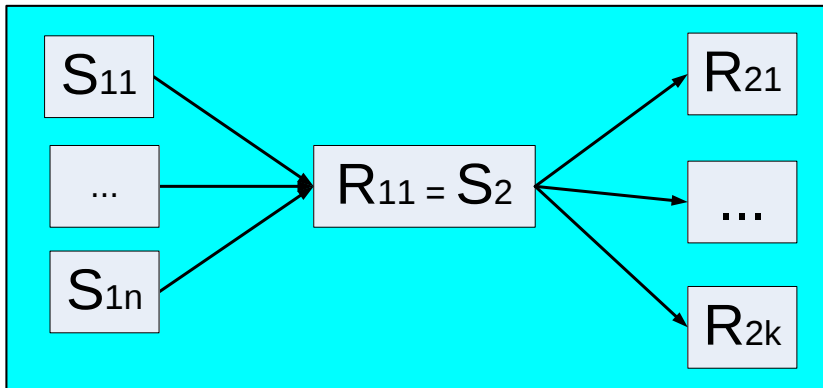
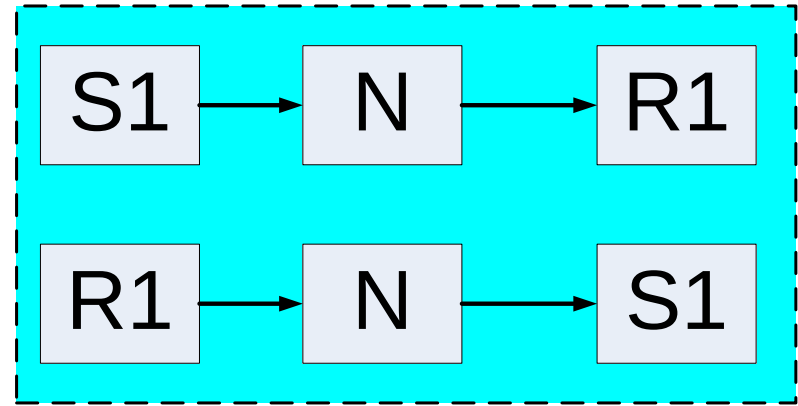
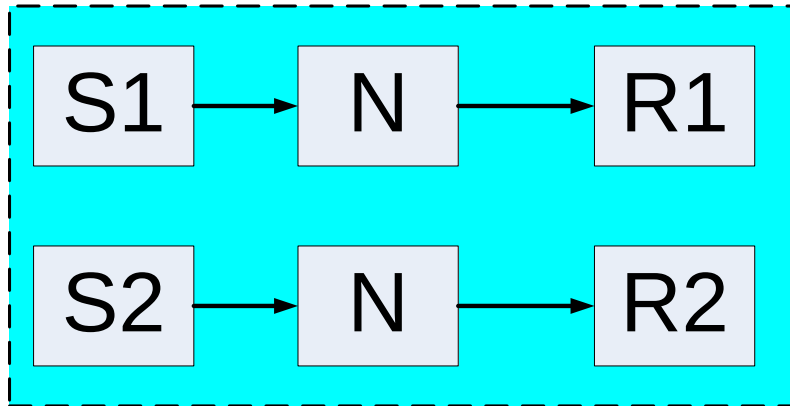
$$S_1 \rightarrow \{(R_i = S_i)\} \rightarrow \dots \rightarrow \{R\}$$

Associationism vs. Connectionism



- ▶ Hebb's associative mechanism

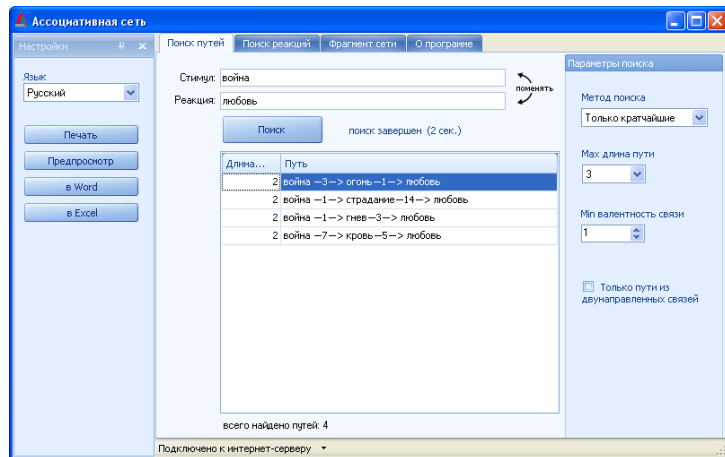
Associative neuron and memory



$$\omega_{S_{11}R_{11}} = \frac{f_{omH}(S_{11} \rightarrow R_{11}) * K}{\sum_{i=1}^n f_{omH}(S_{1i} \rightarrow R_{1i})}$$

- Study structure of AVN
- Revealing «metaassociations»
- Analysis derivative associations

Program modules for searching associative chains



- ▶ **Alexander Panchenko, 2008**
The program makes possible multifarious kinds of search for chains (AVC) in the associative networks (Russian, English).
- ▶ The program is written in C# and make use of SQL Server DBMS.

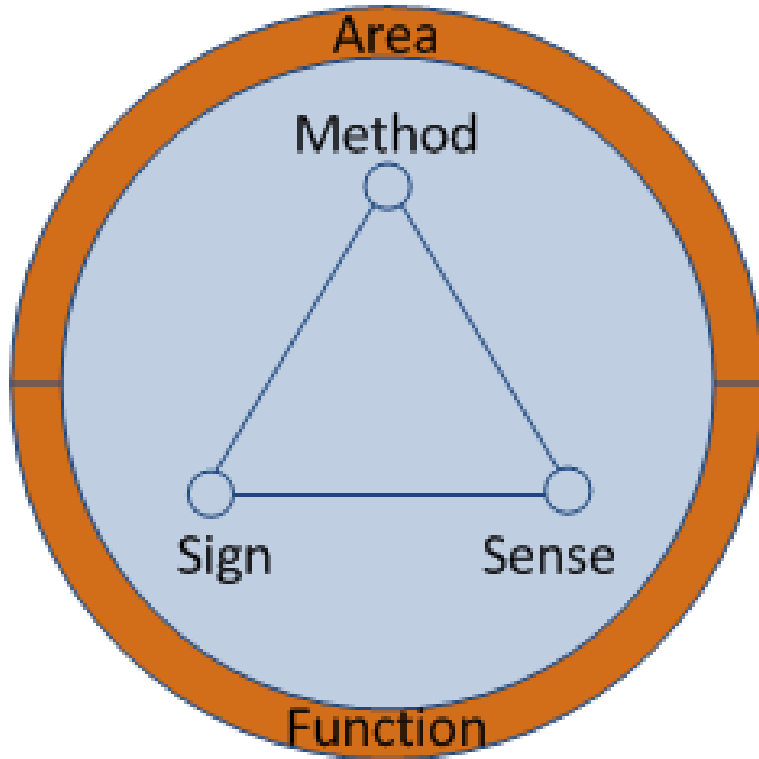


- ▶ **Alexander Sirenko, 2008–2010**
- ▶ Search chains (AVC) in Russian AVN (120 000 associations).
- ▶ AVS is stored in MySQL, data processing with PHP5.
- ▶ <http://thesaurus.ru/aweb.php>

LINGUACULTURAL THESAURI



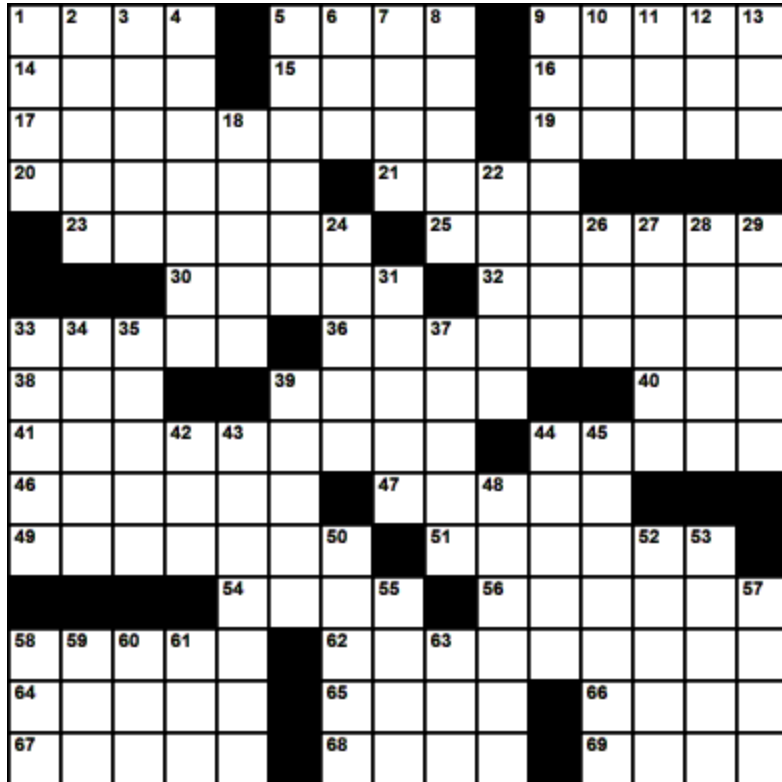
Figure of knowledge – Cognem



- ▶ In the conception of J. N. Karaulov **Cognem** — the elementary unit of knowledge:
 - Sign (word)
 - Sense (semantics)
 - Methods (for define sense) 40+
 - Two type of knowledge functions:
 - Recipe
 - Retouch

- ▶ Two modes of language consciousness activity (Cognizer):
 - **Active (Sign → Sense)**
 - **Passive (Sense → Sign)**

Source of cognems –crosswords



Across

- 1. A 1977 movie
- 5. Actor Garcia
- 9. City in the Texas panhandle
- 14. 82nd element
- 15. Carpooled
- 16. Mucous eye discharge
- 17. Quadrennial prize
- 19. Ahead of schedule
- 20. Irish form of Helen

Example of cognems

<i>Formula of sense</i>	<i>Area</i>	<i>Function</i>	<i>Method</i>
Trailer truck	Machines(Cars)	Recipe	Synonym
A male cat	Animals	Retouch	Mark
Self-satisfied	The feature of the human	Recipe	Definition
A city in east-central France	Cities	Retouch	Mark
A place for wives and concubines	Constructions	Retouch	Description
Protagonist	Narration	Retouch	Synonym
Formerly (archaic)	Indexes of time	Retouch	Synonym
Sporting venue	Sports	Recipe	Perifraz
The boundary of a surface	Objects	Recipe	Synonym
Rodent	Rodents	Recipe	Hyperonym-hyponym
Blue-green	Color	Recipe	Synonym
What doors swing on	Building designs	Retouch	Description
Swerved	Travel	Recipe	Synonym
Long period of time	Indexes of time	Retouch	Description

Thesaurus article

- ▶ **Can.** *Recipe.* milk container. <milk; container> → <can> = <CAN
1 → CONTAINER 1 → MILK; CAN 1 → SOUP 1 → MILK; CAN 1 → TEA
2 → MILK; CAN 1 → CONTAINER; > 4=1+3+0+0+0

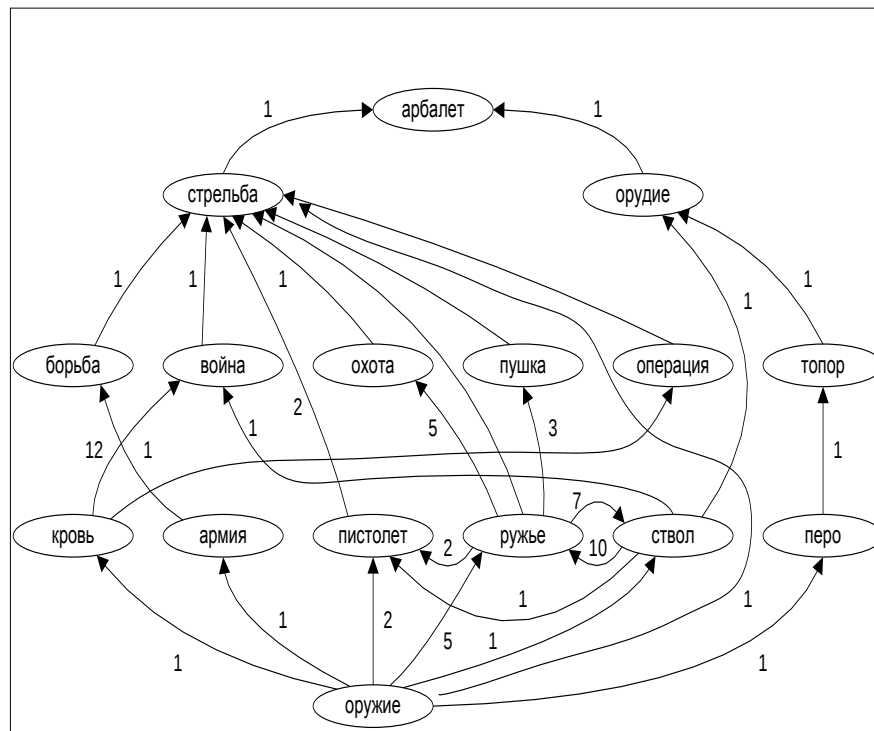
- ▶
- ▶ <MILK 1 → COCOA 1 0 → BEANS 1 → CAN; MILK 5 → BOTTLE 1 7 →
BEER 1 → CAN; MILK 1 1 → DRINK 9 → BEER 1 → CAN; MILK 1 →
HEALTH 1 → BEER 1 → CAN; MILK 1 → JAR 3 → BEER 1 → CAN; MILK
1 → OFF 1 → BEER 1 → CAN; MILK 1 → PUDDING 1 → BOWL 1 →
CAN; MILK 1 → SICK 1 → BOWL 1 → CAN; MILK 1 1 → DRINK 2 →
COCA-COLA 1 → CAN; MILK 1 → COOL 1 → COKE 1 → CAN; MILK
1 → JAR 4 → CONTAINER 2 → CAN; MILK 3 → BABY 1 → DON'T 1 →
CAN; MILK 1 → JAR 3 → LID 3 → CAN; MILK 1 → TOP 1 → LID 3 →
CAN; MILK 1 → CALCIUM 1 → METAL 1 → CAN; MILK 1 → BUTTER
1 → OIL 2 → CAN; MILK 1 → LARD 1 → OIL 2 → CAN; MILK 3 →
MACHINE 2 → OIL 2 → CAN; MILK 5 → BOTTLE 1 → OPENER 3 4 →
CAN; MILK 1 → COCOA 3 → TIN 3 1 → CAN; CONTAINER 2 → CAN;
> 21=1+0+20+0+0

Thesaurus article

- ▶ **Can. Recipe.** milk container. <milk; container> → <can> = <CAN
 $1 \rightarrow \text{CONTAINER}$ $1 \rightarrow \text{MILK}$; CAN $1 \rightarrow \text{SOUP}$ $1 \rightarrow \text{MILK}$; CAN $1 \rightarrow \text{TEA}$
 $2 \rightarrow \text{MILK}$; CAN $1 \rightarrow \text{CONTAINER}$; > **4=1+3+0+0+0**
- ▶
- ▶ <MILK $1 \rightarrow \text{COCOA}$ $1^0 \rightarrow \text{BEANS}$ $1 \rightarrow \text{CAN}$; MILK $5 \rightarrow \text{BOTTLE}$ $1^7 \rightarrow$
 BEER $1 \rightarrow \text{CAN}$; MILK $1^1 \rightarrow \text{DRINK}$ $9 \rightarrow \text{BEER}$ $1 \rightarrow \text{CAN}$; MILK $1 \rightarrow$
 HEALTH $1 \rightarrow \text{BEER}$ $1 \rightarrow \text{CAN}$; MILK $1 \rightarrow \text{JAR}$ $3 \rightarrow \text{BEER}$ $1 \rightarrow \text{CAN}$; MILK
 $1 \rightarrow \text{OFF}$ $1 \rightarrow \text{BEER}$ $1 \rightarrow \text{CAN}$; MILK $1 \rightarrow \text{PUDDING}$ $1 \rightarrow \text{BOWL}$ $1 \rightarrow$
 CAN; MILK $1 \rightarrow \text{SICK}$ $1 \rightarrow \text{BOWL}$ $1 \rightarrow \text{CAN}$; MILK $1^1 \rightarrow \text{DRINK}$ $2 \rightarrow$
 COCA-COLA $1 \rightarrow \text{CAN}$; MILK $1 \rightarrow \text{COOL}$ $1 \rightarrow \text{COKE}$ $1 \rightarrow \text{CAN}$; MILK
 $1 \rightarrow \text{JAR}$ $4 \rightarrow \text{CONTAINER}$ $2 \rightarrow \text{CAN}$; MILK $3 \rightarrow \text{BABY}$ $1 \rightarrow \text{DON'T}$ $1 \rightarrow$
 CAN; MILK $1 \rightarrow \text{JAR}$ $3 \rightarrow \text{LID}$ $3 \rightarrow \text{CAN}$; MILK $1 \rightarrow \text{TOP}$ $1 \rightarrow \text{LID}$ $3 \rightarrow$
 CAN; MILK $1 \rightarrow \text{CALCIUM}$ $1 \rightarrow \text{METAL}$ $1 \rightarrow \text{CAN}$; MILK $1 \rightarrow \text{BUTTER}$
 $1 \rightarrow \text{OIL}$ $2 \rightarrow \text{CAN}$; MILK $1 \rightarrow \text{LARD}$ $1 \rightarrow \text{OIL}$ $2 \rightarrow \text{CAN}$; MILK $3 \rightarrow$
 MACHINE $2 \rightarrow \text{OIL}$ $2 \rightarrow \text{CAN}$; MILK $5 \rightarrow \text{BOTTLE}$ $1 \rightarrow \text{OPENER}$ $3^4 \rightarrow$
 CAN; MILK $1 \rightarrow \text{COCOA}$ $3 \rightarrow \text{TIN}$ $3^1 \rightarrow \text{CAN}$; CONTAINER $2 \rightarrow \text{CAN}$;
 > **21=1+0+20+0+0**

Cognem Dictionary/Thesaurus

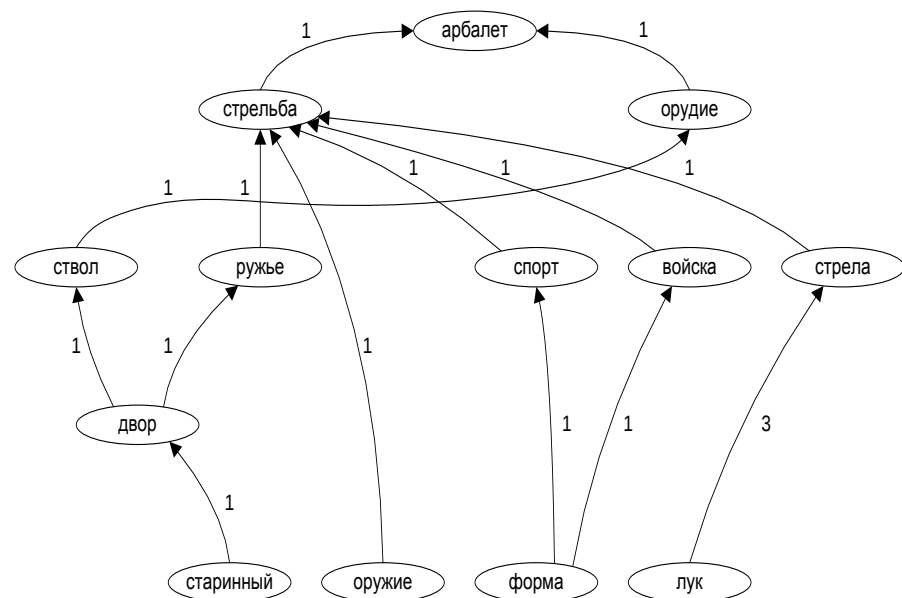
	оружие
1	оружие -1-> стрельба
2	оружие -2-> пистолет -2-> стрельба
2	оружие -5-> ружье -1-> стрельба
3	оружие -5-> ружье -5-> охота -1-> стрельба
...	
3	оружие -1-> армия -1-> оружие -1-> стрельба
4	оружие -1-> армия -1-> мучение -1-> борьба -1-> стрельба
...	
4	оружие -1-> ствол -1-> пушка -1-> чугун -1-> стрельба
<i>Всего цепочек 141, в том числе: 1 — нет, 2 — 1, 3 — 2, 4 — 15, 5 — 123.</i>	
2	оружие -1-> ствол -1-> орудие
3	оружие -5-> ружье -1-> ствол -1-> орудие
3	оружие -1-> перо -1-> топор -1-> орудие
4	оружие -2-> пистолет -1-> железо -3-> лом -1-> орудие
...	
4	оружие -1-> армия -28-> солдат -1-> топор -1-> орудие
<i>Всего цепочек 29, в том числе: 1 — нет, 2 — нет, 3 — 1, 4 — 2, 5 — 26</i>	



Passive mode of cognizer

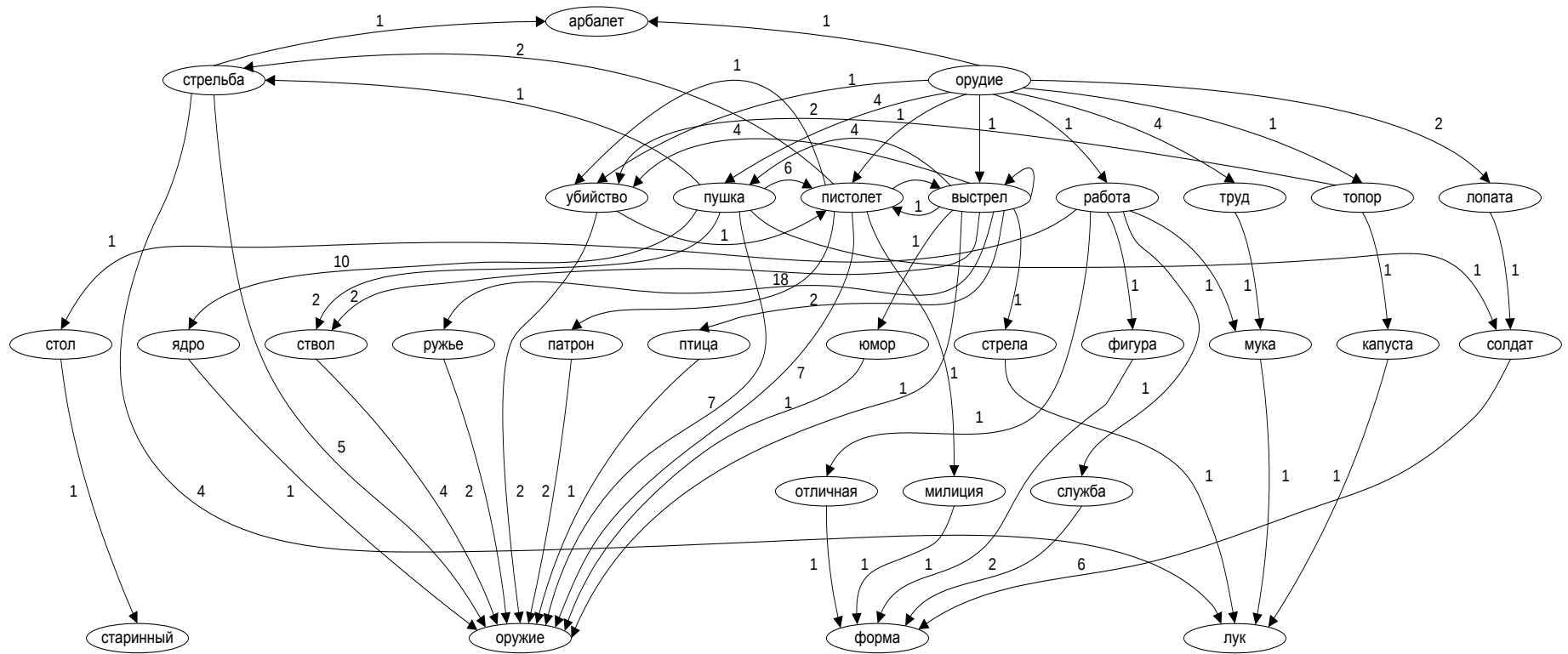
цепочки	длина цепочки					всего
	1	2	3	4	5	
старинный...→стрельба→арбалет	нет	нет	нет	1	30	31
оружие...→стрельба→арбалет	нет	1	2	15	123	141
форма...→стрельба→арбалет	нет	нет	2	4	98	104
лук...→стрельба→арбалет	нет	нет	1	нет	20	21
всего	нет	1	5	20	271	297
старинный...→орудие→арбалет	нет	нет	нет	1	21	22
оружие...→орудие→арбалет	нет	нет	1	2	26	29
форма...→орудие→арбалет	нет	нет	нет	1	9	10
лук...→орудие→арбалет	нет	нет	нет	2	2	4
всего	нет	нет	1	6	58	65
всего	нет	1	6	26	329	362

4	старинный —1→ двор —1→ ствол —1→ орудие —1→ арбалет
4	старинный —1→ двор —1→ ружье —1→ стрельба —1→ арбалет
2	оружие —1→ стрельба —1→ арбалет
3	форма —1→ спорт —1→ стрельба —1→ арбалет
3	форма —1→ война —1→ стрельба —1→ арбалет
3	лук —3→ стрела —1→ стрельба —1→ арбалет
Всего цепочек 6, в том числе: 1 — нет, 2 — 1, 3 — 3, 4 — 2, 5 — нет	



Minimal passive proposition graph

Active mode of Cognizer



Active proposition graph (dimension 4)