

4th Annual

# Global TechMining Conference

2014.09.02 • Leiden, Netherlands

## Peaks, Slopes, Canyons, Plateaus: Identifying Technology Trends throughout the Life Cycle

Irina Efimenko, Vladimir Khoroshevsky,  
Oleg Ena

# Technology Trend Enigma

- ▶ Complex nature, low formalization level, blurred boundaries, high degree of domain dependency
  - ▶ the need for expert knowledge
- ▶ Too general names (clean energy, semantic technologies, etc.)
- ▶ *△: an illumination lightpipe with a dot pattern to redistribute the collected sunlight from the natural light illumination system – ?*

## However

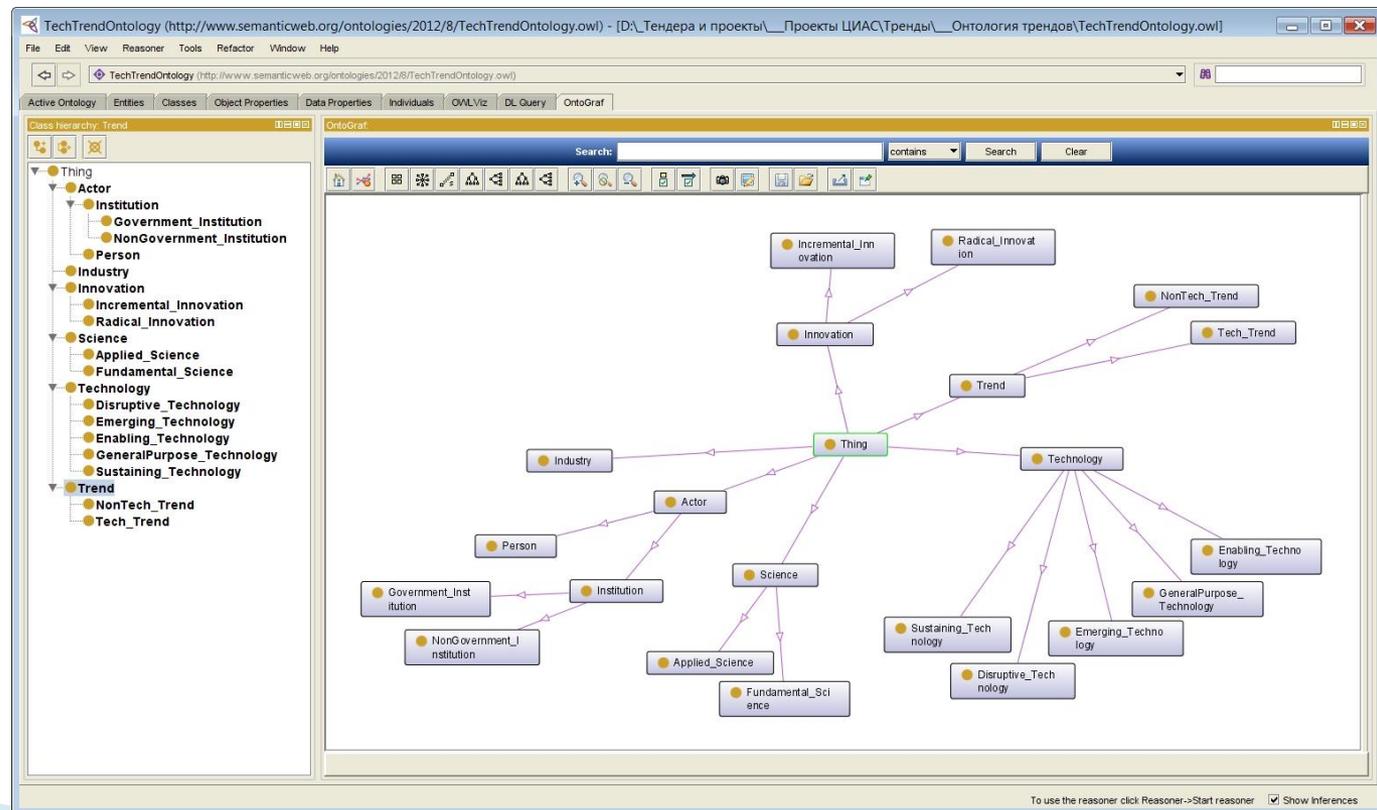
- ▶ We name both 'Big Data' in IT and 'Genome Editing' in Healthcare 'a TT'.

# Formal Model-1: BB approach



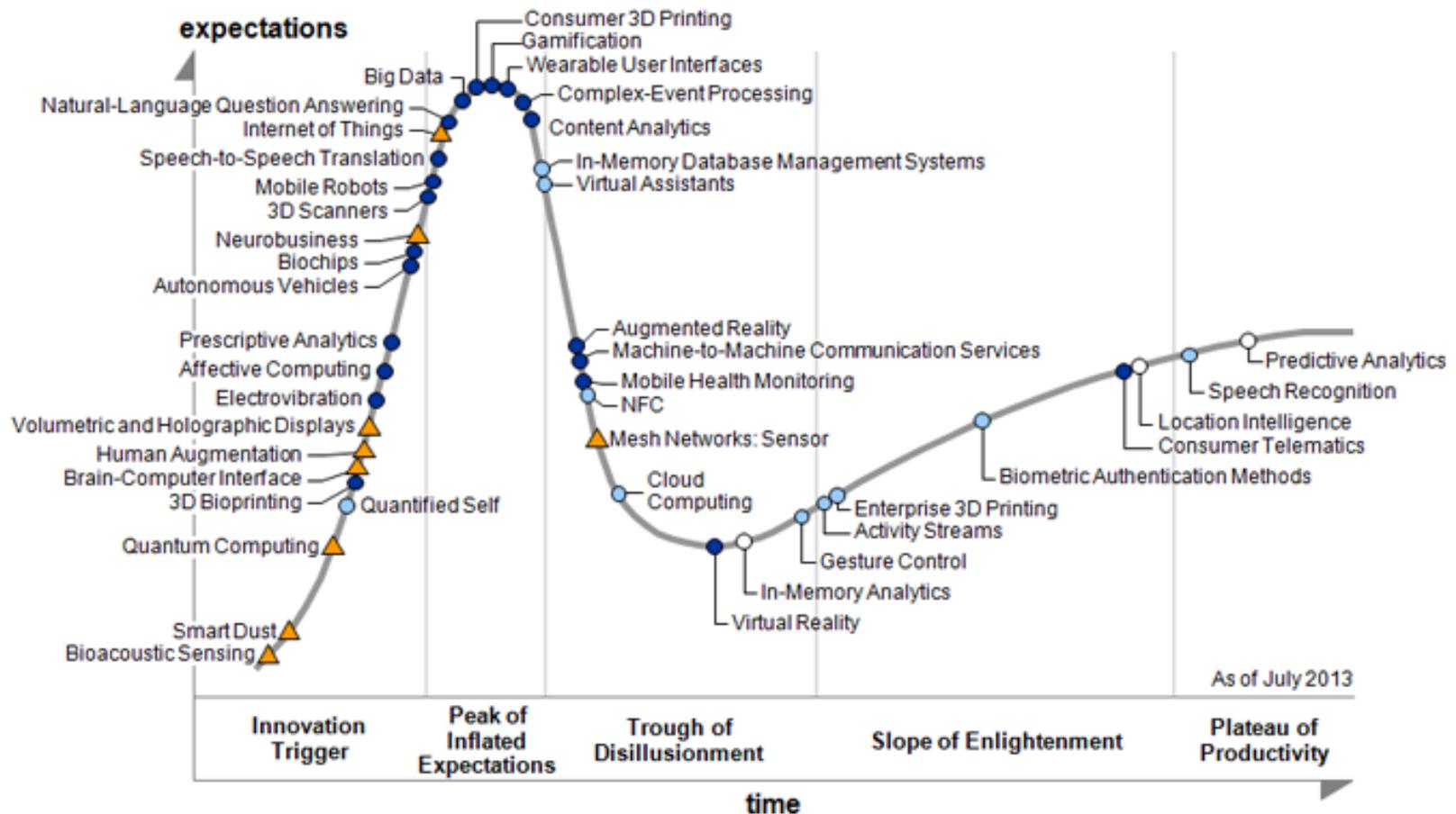
# Formal Model-2

- ▶ Static component: Ontology of a TT for specification of indicators



# Formal Model-3

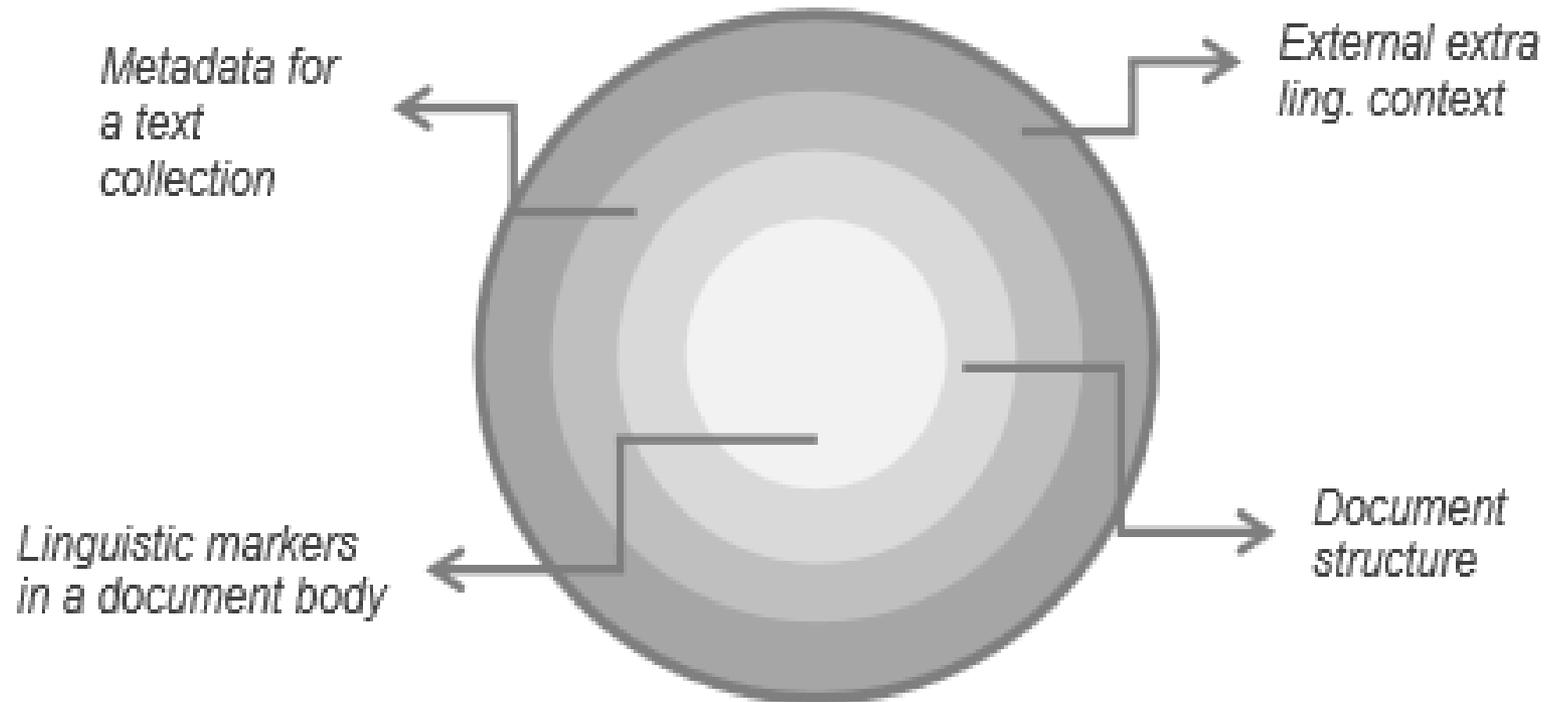
- ▶ Dynamic component: Gartner's Technology Hype Cycle



# Life Cycle: Basic Assumptions

- ▶ Specific (extra) linguistic markers should be used for each phase (indicators from ontology are mapped onto the HTLC);
- ▶ TT monitoring should be carried out based on a mixture of text genres. Each phase has its own most relevant genres
- ▶ Patent analysis is most important for the phases of Slope of Enlightenment and Plateau of Productivity
- ▶ Surge and loss of interest could be fixed based on processing of technology news and specific types of foresight and analytical reports. It is relevant for the Peak of Inflated Expectations and the Trough of Disillusionment
- ▶ Scientific papers and R&D reports are relevant for the Technology Trigger. It seems to be the most interesting phase for the TT monitoring task (at the stage of **weak signals**).

# Heuristics and Indicators



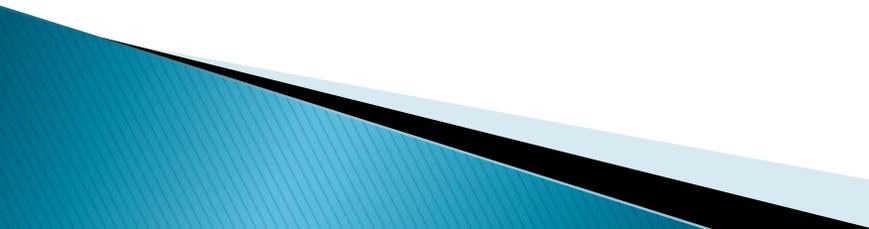
# Collection Level and Extra linguistic Indicators

- ▶ Dynamics of co-citation in different shadow groups and virtual teams of authors
  - ▶ New sections at scientific conferences or in scientific journals
  - ▶ New awards
  - ▶ Calls for projects
  - ▶ Specific phenomena in term usage (e. g. geographical or institutional expansion)
- 

# Structural Indicators of Text Relevance

- ▶ Abstracts
- ▶ Lists in reports
  - *△ At MEPC 62 the Institution was involved in discussions on:*
    - *Energy Efficiency Design Index (EEDI);*
    - *Prevention of air pollution;*
    - *Reduction in Green House Gases...*

Etc.



# Linguistic Markers and Events–1

- ▶ Needs and Values (health, safety, stability, etc.) – *No is less than for solutions*
- ▶ Resources (natural, financial, human resources, time, etc.)
- ▶ Technological problems; Technological and economic parameters (price, size, safety, energy efficiency, etc.)
- ▶ Markers of scientific or technological process / results (experiments, model, method, production, market, technology, etc.)
- ▶ New types of objects and their comparison (*X type of Y, like most X-s, X-s are a category of Y that..., X solution*)

# Linguistic Markers and Events–2

Linguistic Scales indicate the mentioning of

- (a) new / existing needs (customers' pains)
- (b) resources, values
- (c) S&T issues (tasks)
- (d) technological parameters (the change of).

They often indicate “the most important achievement” of a new technology proposed in a paper / being analyzed in a report, etc.

# Linguistic Markers and Events–3

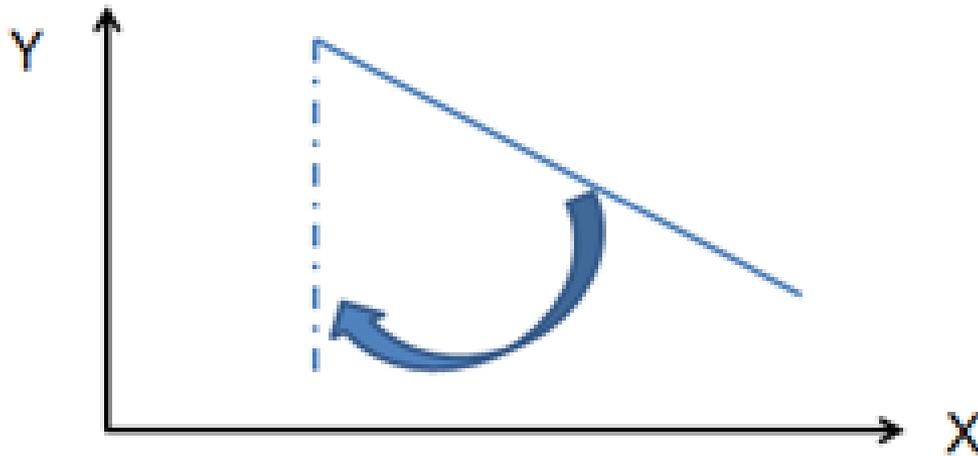


- ✓ The US Export–Import Bank is to sign a \$2 billion deal with South Africa to fund a green energy scheme in the electricity–short country.
- ✓ Defects that are of zero or very small volume, known as kissing defects, are *much* harder to locate.
- ✓ There is *considerably* less understanding about the performance of IBs.
- ✓ *Remarkable* reduction in tensile, flexural and interlaminar properties was noticed after 2 weeks of immersion for all three materials.
- ✓ When increasing the external magnetic field by only 50 mTesla the dynamic stiffness for isotropic samples increased by 100 % while the damping factor decreased by 17%.
- ✓ For the marine industry where unpredictable dynamic loading conditions are the case, MRE isolators could *greatly* decrease the level of vibrations transmitted from the machines to the shell of the ship and the opposite, resulting to smaller fatigue loads and a *much* more comfortable journey

# Linguistic Markers and Events-4

- ▶ “Special” events:
  - Decrease of correlation for interrelated parameters (for inverse dependence)

Potential evidence for a disruptive technology at the stage of weak signals



# Linguistic Markers and Events-5

- ▶ The fact of forming a new regulatory committee.

Potential evidence for a more mature TT

*the need for regulation means that the phenomenon is newish, but already having significant impact*

# Markers as Stop Words

- ▶ Identification of the most relevant text fragments for further processing (key terms extraction)
  - ▶ Identification of a Life Cycle stage (experiments, production, etc.)
  - ▶ Semantic “Stop-Fields” (they themselves are not included into the output)
  - ▶ “Global Context” (not just TTs, but policy making, economics, environmental aspects, etc.)
- 

# General Pipeline

- ▶ Hybrid linguistic + statistical approach
- ▶ Own software for NL text processing combined with existing one for visualization, bibliometric analysis, etc.

GATE Developer 7.0 build 4195

File Options Tools Help

GATE

Applications

- SaveApps
- English RoadMap
- ANNIE
- RusRoadmapIE

Language Resources

- wos132472.txt...99.txt...449...2002.txt.00315
- wos132472.txt...98.txt...36550...2002.txt.00314
- wos132472.txt...98.txt...448...2002.txt.00313
- wos132472.txt...98.txt...36549...2002.txt.00312
- wos132472.txt...97.txt...447...2002.txt.00311
- wos132472.txt...97.txt...36548...2002.txt.00310
- wos132472.txt...96.txt...446...2002.txt.0030F
- wos132472.txt...96.txt...36547...2002.txt.0030E
- wos132472.txt...95.txt...445...2002.txt.0030D
- wos132472.txt...95.txt...36546...2002.txt.0030C
- wos132472.txt...94.txt...444...2002.txt.0030B
- wos132472.txt...94.txt...36545...2002.txt.0030A
- wos132472.txt...93.txt...443...2002.txt.00309
- wos132472.txt...93.txt...36544...2002.txt.00308

MatchesAnnots [null=[[533, 535]]]

MimeType text/plain

docNewLineType

gateSourceURL file:/D:/CorpusPureAnnotationWithYear/2002/6000/

Messages English RoadMap CorpusEng wos132472.txt...

Annotation Sets Annotations List Annotations Stack Co-reference Editor Text

In the fabrication of W-Cu nanocomposite powders, it is important to consider the design of the alloying content and microstructure to get optimal thermal properties. However, W and Cu have mutual insolubility and their contact angle is very large, thus it is difficult to fabricate homogeneous and fully dense composite materials. Therefore, in recent years studies have been made of the fabrication of W-Cu nanocomposites through the reduction of WO<sub>3</sub>-CuO powders. In the present study, WO<sub>3</sub> and CuO powders were ball-milled with high energy at 400 rpm for 1-50h in Ar atmosphere. They were examined using XRD, SEM and TEM. After the sieved WO<sub>3</sub>-CuO powders were reduced in dry hydrogen, the reduced W-Cu powders were liquid-phase sintered at 1200 and 1300degreesC for 1h in dry hydrogen. After ball-milling of 1h, the reduced powder had a torred structure and the W particles were faceted. However, after ball-milling of 20h, the powder had a homogeneously mixed structure and the W particles were smaller than from the reduced powder ball-milled for 1h. After sintering at 1200degreesC, the W particles were faceted and their microstructure with 30% of well-dispersed and a distribution of grain sizes of 100-200nm.

Type	Set	Start	End	Id	Features
KeyExpression		7	48	599	{chunkNorm=fabrication of W-Cu nanocomposite powders}
Keyword		7	18	577	{chunkNorm=fabrication}
KeyExpression		22	48	578	{chunkNorm=W-Cu nanocomposite powders}
Keyword		82	88	579	{chunkNorm=design}
KeyExpression		96	131	573	{chunkNorm=alloying content and microstructure}
KeyExpression		139	165	574	{chunkNorm=optimal thermal properties}
KeyExpression		176	184	575	{chunkNorm=W and Cu}
KeyExpression		190	209	576	{chunkNorm=mutual insolubility}
KeyExpression		220	233	570	{chunkNorm=contact angle}
KeyExpression		283	330	571	{chunkNorm=homogeneous and fully dense composite materials}
KeyExpression		346	366	572	{chunkNorm=recent years studies}
KeyExpression		389	423	600	{chunkNorm=fabrication of W-Cu nanocomposites}
Keyword		389	400	553	{chunkNorm=fabrication}
KeyExpression		404	423	554	{chunkNorm=W-Cu nanocomposites}
KeyExpression		436	452	601	{chunkNorm=reduction of WO3}
Keyword		436	445	555	{chunkNorm=reduction}
KeyExpression		449	452	556	{chunkNorm=WO3}
KeyExpression		453	464	552	{chunkNorm=CuO powders}
KeyExpression		473	486	569	{chunkNorm=present study}
KeyExpression		488	491	567	{chunkNorm=WO3}
KeyExpression		496	507	568	{chunkNorm=CuO powders}
KeyExpression		530	541	565	{chunkNorm=high energy}
KeyExpression		545	552	566	{chunkNorm=400.rmm}

48 Annotations (0 selected) Select:

Document Editor Initialisation Parameters

Views built

Annotations Stack

- Date
- Expr
- FirstPerson
- Identifier
- KeyExpression
- Keyword
- Lookup
- NounChunk
- Person
- Sentence
- SpaceToken
- Split
- Token
- Unknown
- VG
- Word

Original markups

# Case Study: Clean Energy

- ▶ Text collections in Clean Energy domain (131,477 documents in English and in Russian, scientific papers, technology news, etc.; 2002–2014)
- ▶ Examples:
  - WoS (2008–2013)
  - Technology news (mostly got from <http://cleanedge.com/>)
  - Patent analysis is actually a different business task

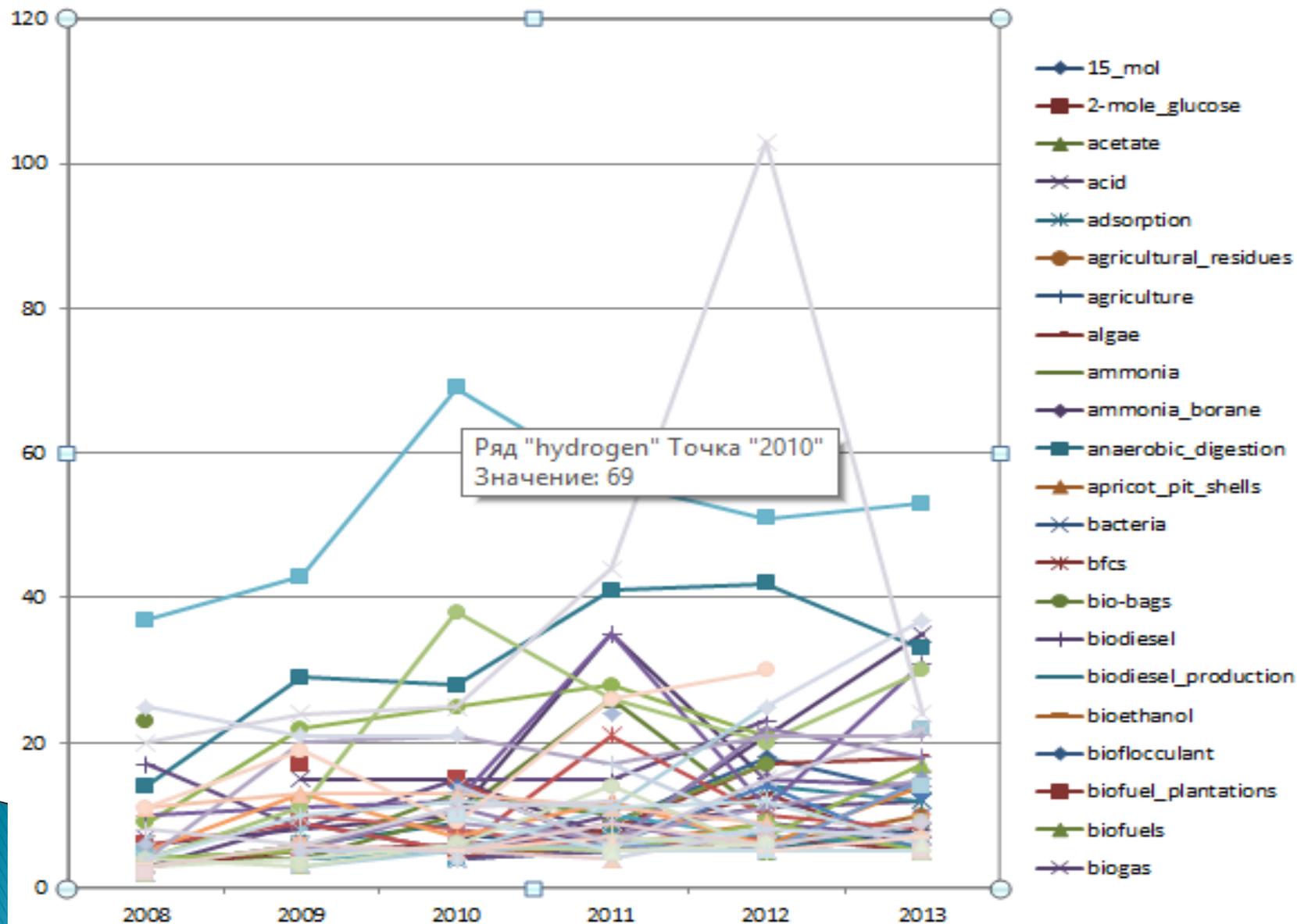
# Scientific Papers: Basic Assumptions

Linguistic indicators are used to process text fields (full texts, abstracts, keywords, titles)

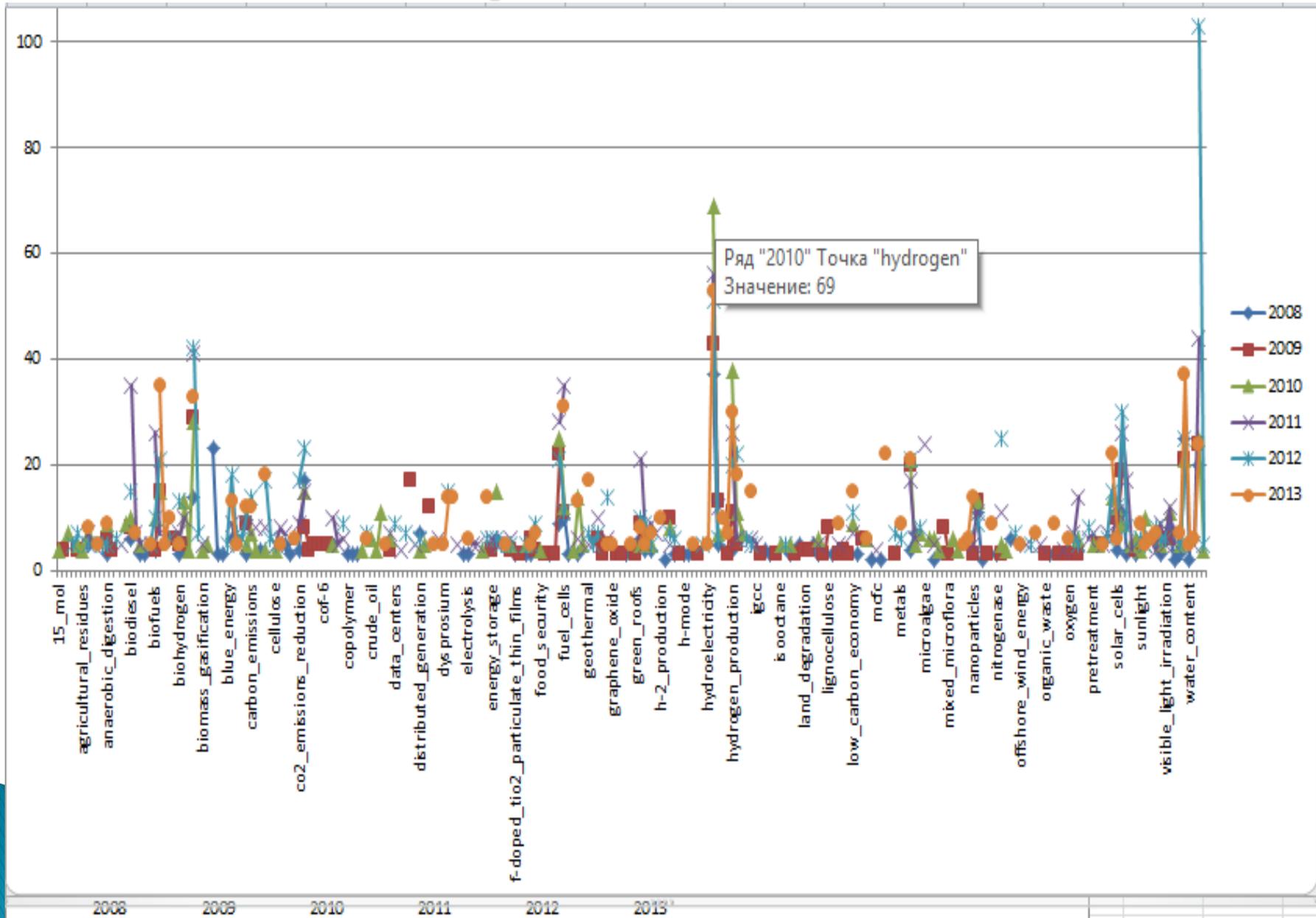
Extra linguistic indicators are used for analytical assumptions:

- ▶ Authors' affiliation: Research Institutes / Universities vs. Companies ▶ more mature stage of R&D
- ▶ Source of a paper: Journal vs. Conference ▶ “hot topic” (stage of Innovation Trigger)
- ▶ Science: Hard Sciences vs. SSH ▶ technology push and market pull
- ▶ Key terms from highly cited papers ▶ more mature + novel solution

# Scientific Papers: Results



# Scientific Papers: Results



# Scientific Papers: Results (2008): Journals

key
biosolids
sewage_sludge
wood_ashes
exergy
dams
ocean_acidification
combustion
a_clean_energy_carrier
land_application
large_dams
hydrogen_energy
energy_recovery
biohydrogen_production
potable_water

key
industrial_so2_emissions
hydrogen_and_fuel_cells
waste-water
willingness-to-pay
f-doped_tio2_particulate_thin_films
feedstock
electrolysis
copper
corn
biofloculant

# Scientific Papers: Results (2008): Conferences

key
Wind
biomass
climate_change
solar_energy
natural_gas
fuel_cells
hydrogen_production
fossil_fuels
waste
distributed_generation
methanol
ethanol
biodiesel
carbon
pemfc

key
greenhouse_gas_emissions
oxidation
cardoon
solar_cells
photovoltaics
methane
the_grid
wastewater
wind_turbine
thermomechanical_units

# Scientific Papers: Results (2013): Journals

key
biomass
catalysts
electrocatalysts
membranes
h-2_production
hydrogen_evolution
li-ion_batteries
agriculture
nanotubes
olivine
tio2/cuo_composite_nanofibers
mos2
zeolitic_imidazolate
wind_turbines
tio2

key
pig_manure
crude_glycerol
carbon_dioxide_emission
composite
co2_capture
ghg_emissions
water_splitting
the_monopile_structure
nanowires
an_rwps

# Scientific Papers: Results (2013): Conferences /comp. to GTLC parameters/

key
climate_change
solar_energy
Biogas
fossil_fuels
Methane
wind_energy
co2_emissions
fuel_cells
wind_power
energy_consumption
Wind
hydrogen_storage
energy_sources
Biofuels
coal

key
smart_grid
hydropower
hydrogen-production
power_generation
electric_vehicles
global_warming
energy_harvesting
gasification
photovoltaics
carbon_emissions

# Scientific Papers: Results (2013): Companies vs. Res. Inst.

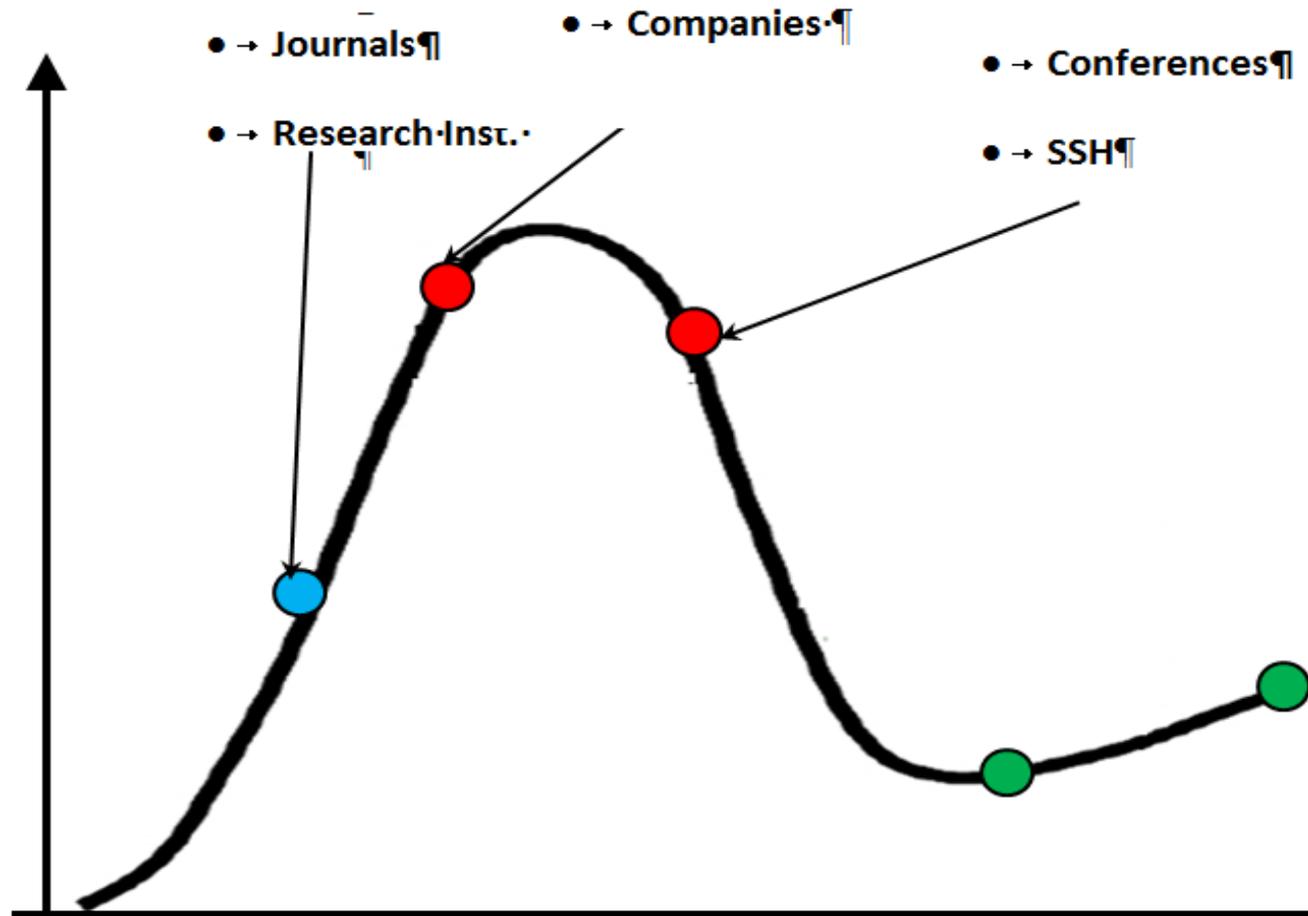
molecular_hydrogen	pem_electrolysis	f-doped_tio2_particulate_thin_films
active_and_passive_material_elements	dampeg	feedstock
lignin	supercapacitors	electrodeposition
recalcitrance	sensors	environmental_taxes
optical-properties	sensitized_solar-cells	clean_energy_conversion
organic_waste	dysprosium	copper
switchgrass	ceo2	corn
stand_systems	green_jobs	bioflocculant
cradle	graphene_oxide	bioethanol
shale_gas	home-made_zeolite	biohydrogen
skutterudites	ethylene	birch
dissolution	the_tripod_structure	biosolids_management
cellular_networks	water_content	carbon_dioxide
co2_sequestration	supergrid	anaerobic_digestion
green_energy_applications	fossil_fuel	seawater
households	fuel-cells	organic_wastewater
energy-transfer	microorganisms	pahs
the_recirc_systems	methanol	perylene_bisimide_derivatives
virtual_power_plants	interfacial_polymerization	luv's
thermal_resources	lithium-ion_batteries	liquor
forward_osmosis	offshore_wind_farms	light
materials_selection	oxynitrides	kinetics
mixtures	yuma	nitrogenase
molecular-dynamics	xenotime	h2o/la/
mineralisation	pla/starch_boxes	green_industrialization

# Scientific Papers: Results (2013): SSH vs. HS

- ▶ climate\_change
- ▶ solar\_energy
- ▶ fossil\_fuels
- ▶ wind\_energy
- ▶ co2\_emissions
- ▶ renewable\_energy\_sources
- ▶ biofuels
- ▶ coal
- ▶ global\_warming
- ▶ photovoltaics
- ▶ natural\_gas
- ▶ energy\_security
- ▶ environmental\_protection
- ▶ renewable\_resources
- ▶ biofuel
- ▶ waste
- ▶ heat
- ▶ ghg\_emissions
- ▶ raw\_materials
- ▶ wind\_turbine
- ▶ climate\_policy
- ▶ feed-in\_tariff
- ▶ oil\_prices
- ▶ low-carbon\_technologies
- ▶ deployment\_policies

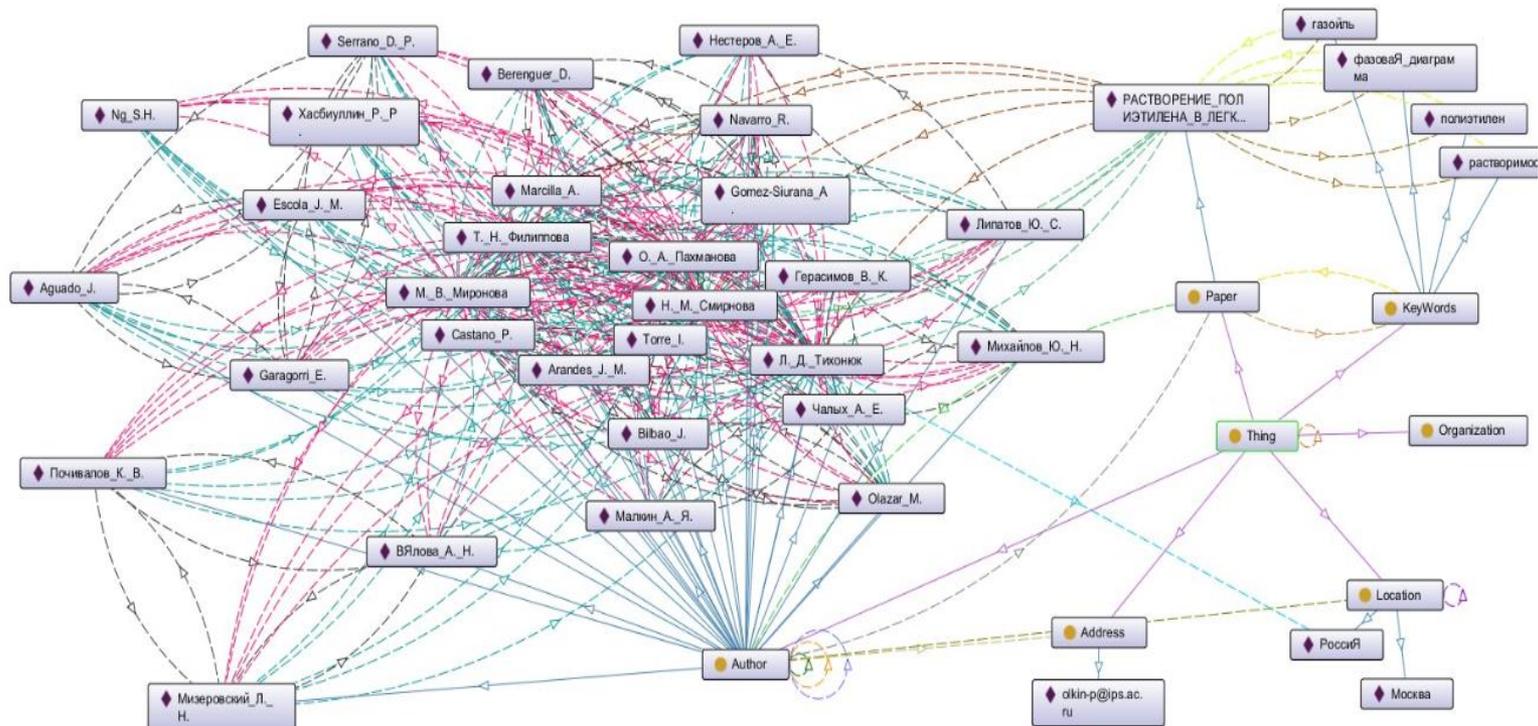
# Scientific Papers: Model

▶ 30+ to ~100



# Scientific Papers: Under Development

- ▶ Geo-expansion; Shadow groups; Research fronts



# Techno News: Results-1 a

2003-2007: core concepts of terms

anti baseline **birds** consultation coordination credit **emissions** extension  
form **gasses** generation **greenhouse** grid handful  
**homes** hurdle **influence** interconnection interest island issues jail  
andowners mapping operation others plan populations portfolio potential preparatory  
production provisions road scope senate site terminals thanks thumbs **town**  
**turbine**wind

# Techno News: Results-1 b

(focus + context)

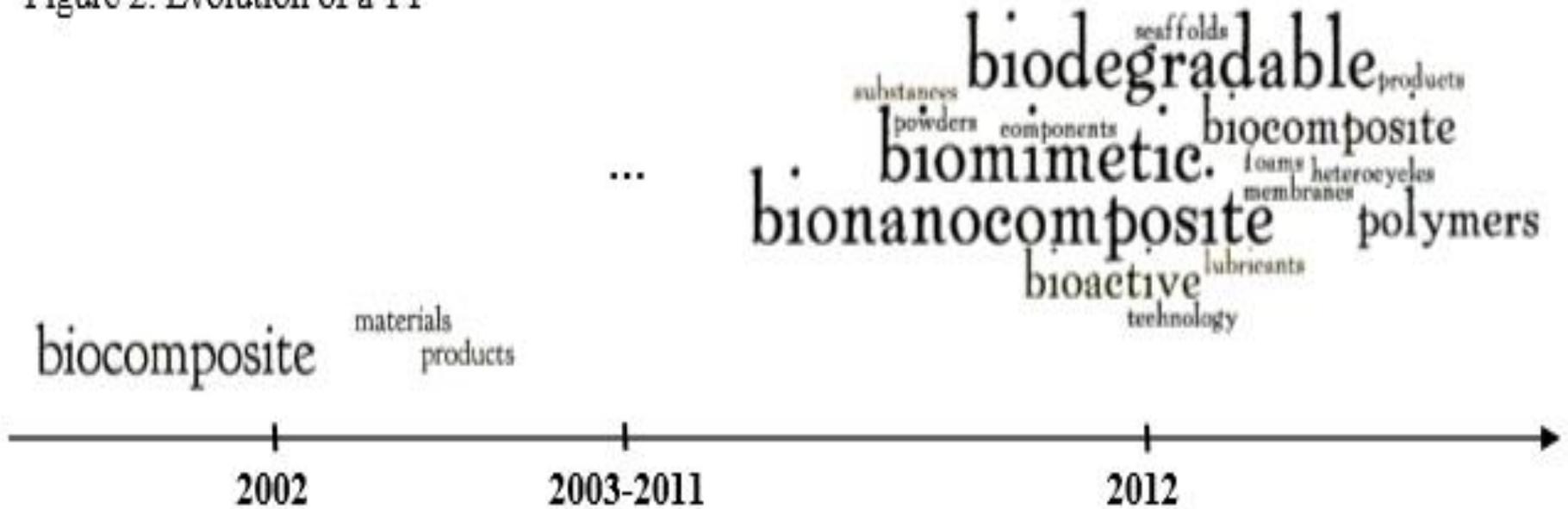
2008-2010: core concepts of terms

aircraft alliance american assessments attention  
automobile **barrels** bill  
**billion** car cells chambers change climate  
combination country dollars economy  
feedstocks **feet** foreign **fuel** gas  
industry infrastructure insulation  
legislation majority manufacturers modernization **oil**  
organizations panel passage plants possibility president protection renjet  
roadmap scrappage sunpower  
transportation **trillion** tropsch  
vehicle warming water winners world

# Techno News: Results-2

## (new heuristics)

Figure 2: Evolution of a TT



# Evaluation

- ▶ Evaluation by experts
- ▶ Comparing to G. Hype Cycle for Emerging Energy Technologies, 2013 and 2014

## Examples

- Entering the Plateau
  - Wind Generation
  - Liquefied Natural Gas

Conclusion: Trend is not a single object: various aspects (views); various stages of the life cycle; various ontological subtypes (ontological nature of the trend – method, concern, approach, innovative products, etc.). Everything depends on the business task

# Further R&D

- ▶ More indicators related to bibliometrics

## *MISSION*

- ▶ Novel methods for visualizing the results (cognitive graphics easily interpretable by decision makers)
  - ▶ Toolkit for analysts
- 

Thank you for your attention!

*Questions?*

[iefimenko@hse.ru](mailto:iefimenko@hse.ru)  
[v.khor@mail.ru](mailto:v.khor@mail.ru)