LEARNING BY ANALOGY IN A HYBRID ONTOLOGICAL NETWORK

Ponomarev S. V.
ponomarev@corp.sputnik.ru
«Sputnik»

Спутник

The Hybrid Ontological Network

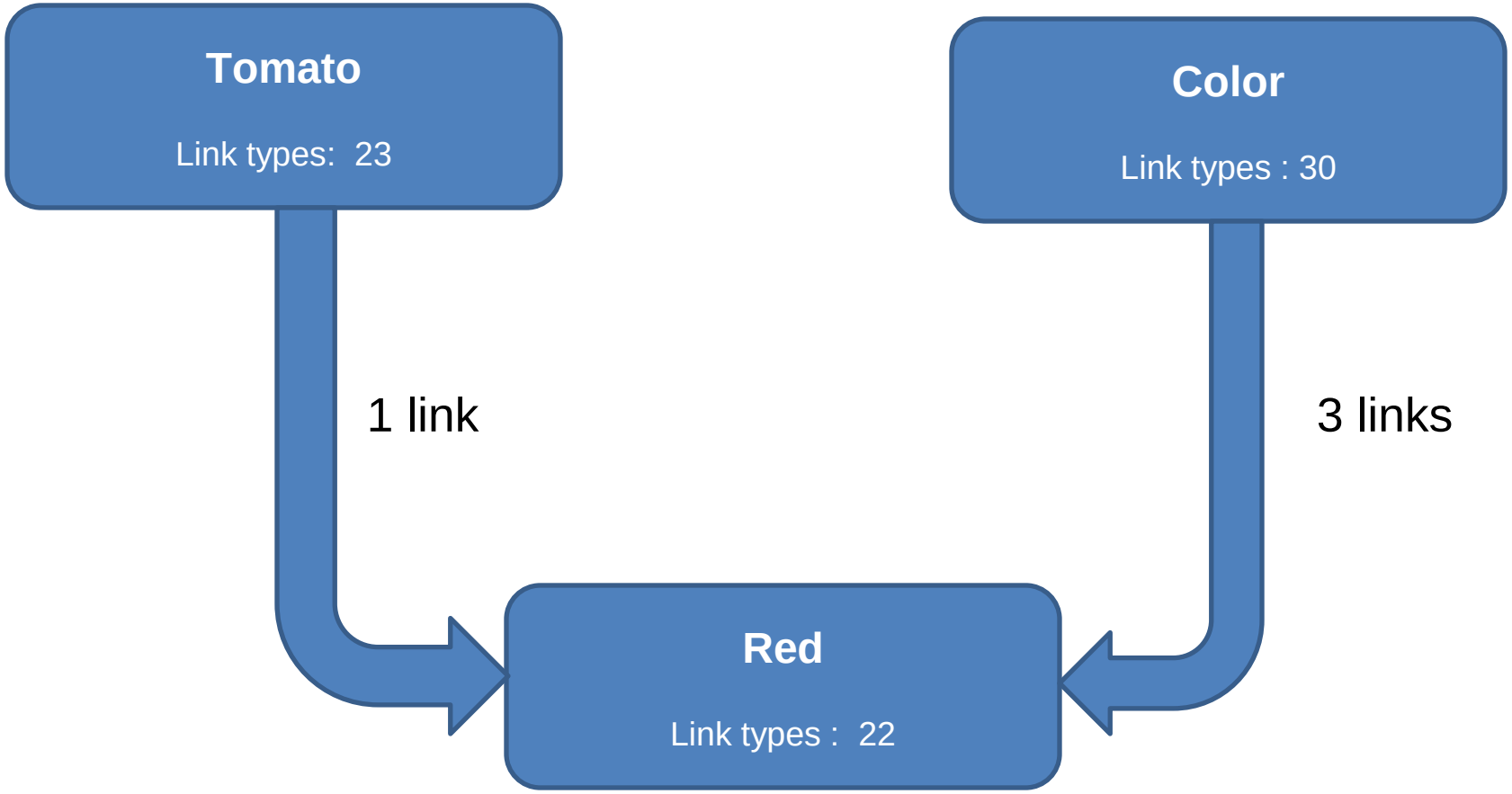
We call an ontology hybrid if:

1. It is composed of several independent sources;
2. It contains triplets (links), accumulated in statistical text corpora processing;
3. Each triplet characterized by type and weight.

The main properties of this ontology are redundancy and high relatedness. Redundancy arises from the duplication of most ontological links in various used sources, and high relatedness arises in inclusion of links obtained by statistical text processing.

Nodes:	1.4 mil.
Links:	639 mil.
Link Types:	64

Network sample



1 * 3 = 3 paths from tomato to color trough red

The link structure between the vertices: "TOMATO", "RED", "COLOR"

	Number of links	Links lead to "RED"
TOMATO	23	1. Related n-grams "adjective + noun", back link (-22).
COLOR	30	1. Hyponyms (8); 2. Related n-grams "adjective + noun", back link (-22); 3. Phrase consists of, back link (-25).

Building a path between the vertices: "TOMATO", "RED", "COLOR"

	Link type	Number of links	Top 10 of the vertices
TOMATO	-22	382	**red**, best, fresh, ripe, rotten, sliced, marinated, rotten, green, salty ...
COLOR	8	7	blue, purple, sea color, orange, **red**, brown, green
COLOR	-22	4032	whole, **red**, white, black, yellow, green, blue, gray, such a, own ...
COLOR	-25	265	versicolour, zinnwaldite, fanal, Black Sea, surah, old gold, cream, blue dust, dark tangerine, light-color ...

Applying constructed path in the network

To build a path in the hybrid network, you should specify a set of pairs "link type – link weight”:

START >> (-22, 1) >> RESULT;

FINISH >> (8, 1/3) >> RESULT;

FINISH >> (-22, 1/3) >> RESULT;

FINISH >> (-25, 1/3) >> RESULT.

START	FINISH	RESULT
TOMATO	COLOR	red - 0.3744 green - 0.238 most - 0.1914 blue - 0.1709 fresh - 0.1691
CURRANT	COLOR	black - 1.1153 red - 1.0118 green - 0.1941 blue - 0.1670 brown - 0.1555
CAR	COLOR	red - 0.1915 own - 0.1789 green - 0.1719 blue - 0.1683 brown - 0.1556
SEA	COLOR	black - 0.2317 red - 0.2293 blue - 0.2233 green - 0.1836 mediterranean - 0.157
SEA	SIZE	length - 0.2732 high - 0.2732 width - 0.25 depth - 0.25 black - 0.2026

The link structure through one intermediate vertice

	All link types	Links lead to "RED"
TOMATO	606	60
COLOR	1018	177
	All link types	**Links lead to "LARGE"**
SEA	1016	107
SIZE	807	108

QA System. Step 1

The rule base is empty, so random response generated. A pair of "question - answer" memorized.

Question: Какой глубины лужа? (What depth is the puddle?)

Correct Answer: Лужа - мелкая. (The puddle is small)

Generated Answer: Глубина. (Depth)

New Rule Added.

QA System. Step 2

In the rule base there’s the only rule obtained in step 1, and the system tries to apply this rule to the question. Attempt fails and the rule is corrected.

Question: Какой глубины море? (What depth is the sea?)

Correct Answer: Море - глубокое. (The sea is deep)

Generated Answer: Море - мелкое. (The sea is small)

Adding 1 New Path.

QA System. Step 3

In the rule base there’s still the only rule, but it’s taught at two examples. The system makes a successful attempt to apply this rule to the question. Thus, in this case two training examples are enough to obtain practically valuable rule.

Question: Какой глубины океан? (What depth is the ocean?)

Correct Answer: Океан - глубокий. (The ocean is deep)

Generated Answer: Океан - глубокий.

(The ocean is deep)Correct Answer Found.

QA System. Step 4

The syntactic structure of pair "question - answer" is changed, so the use of the existing rule does not give the correct result. Another rule is generated.

Question: Какой глубины лужа? (What depth is the puddle?)

Correct Answer: Лужа маленькой глубины. (The puddle is small depth)

Generated Answer: Лужа - мелкая. (The puddle is small)

Generated Answer: Глубина. (Depth)

New Rule Added.

QA System. Step 5

Attempt to apply rule № 2, obtained in step 4, gives the correct result within meaning, but not coinciding exactly with the correct answer. Rule № 2 is corrected.

Question: Какой глубины море? (What depth is the sea?)

Correct Answer: Море большой глубины. (The sea is deep depth)

Generated Answer: Море огромной глубины. (The sea is vast depth)

Adding 1 New Path.

QA System. Step 6

The syntactic structure of pair "question - answer" corresponds more with the rule № 2, than with the rule №1. The attempt to apply rule №2 to determine the color instead of the size gives the expected result.

Question: Какого цвета огурец? (What color is the cucumber?)

Correct Answer: Огурец зеленого цвета. (The cucumber is green color)

Generated Answer: Огурец зеленого цвета. (The cucumber is green color)

Correct Answer Found.

The structure of rule №2

START	Weight of path	Path	RESULT	Weight of path	Path	FINISH
color depth size	0.235216	26 0 7	green red big small	0.168555	3 -7 27 9	cucumber tomato seed sea puddle
	0.117608	25 0 7		0.134844	12 -9 24 -16	
	0.117608	28 0 7		0.134844	-15 -9 24 -16	
	0.0996208	7 -16		0.0374567	3 -3 26 24	
	0.0958917	3		0.0345754	5 -27 -32 24	
	0.0740494	3 27 0 -8		0.0345754	-8 -27 -32 24	
	0.0282505	-25 27 0 -8		0.0313329	-25 15 7	
	0.0270013	6 0 -10 -23		0.0280925	2 -10 -27 5	
	0.0270013	-9 0 -10 -23		0.023249	4 -7 27 9	
	0.0270013	-28 0 -10 -23		0.0210694	3 27 -29 12	

Thank you

Ponomarev S. V.
ponomarev@corp.sputnik.ru
«Sputnik»

Спутник