

ECML PKDD 2009
European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases

<http://www.ecmlpkdd2009.net>
Bled, Slovenia - September 7-11, 2009

NEWSLETTER

W E D N E S D A Y 9 .
S E P T E M B E R

TODAY'S HIGHLIGHTS

- 9:00 - 10:30

Invited talk: Ralf Steinberger

- 16:00 - 24:00

Trip to Postojna and Dinner at the Predjamski dvorec restaurant

WEATHER



We are expecting 21° C/69°F (max), 10°C/50°F (min) in Postojna.

The constant temperature in Postojna Cave is about 10°C/50°F, so be sure to bring along warm clothing.

NEWSLETTER CONTACT

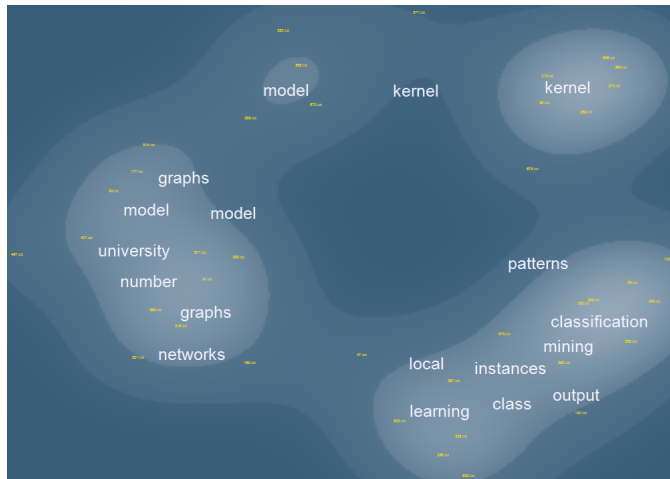
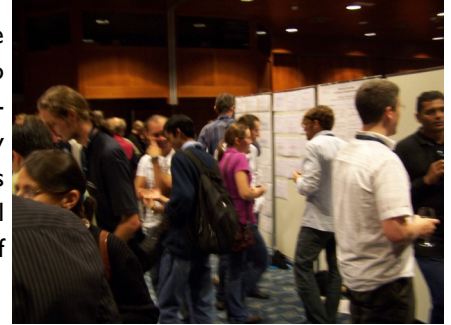
For comments, remarks, questions regarding this newsletter, please email: ecmlpkdd09newsletter@gmail.com

Second Conference Day...

The second conference day was one dedicated to, along with the parallel research tracks, the industry track, the demo and poster sessions.

The “take home” message of Shai Ben-David’s talk was

that there is a large theory-practice gap in the machine learning field, as many learning paradigms used in practice fall beyond the scope of available analysis.



Therefore, the common theoretical models need to be extended.

If you didn't manage to attend all of yesterday's talks, here is what you missed:

clusters, model, usa, kernel, patterns, classification, learning, university, networks, graphs

<http://ontogen.ijs.si/>

Invited Talk: Ralf Steinberger



Today's invited talk will be given by Ralf Steinberger, from European Commission - Joint Research Centre, Italy. His talk is entitled “Highly multilingual news analysis applications”.

ABSTRACT:

The publicly accessible Europe Media Monitor (EMM) family of applications (<http://press.jrc.it/overview.html>) gather and analyse an average of 80,000 to 100,000 online news articles per day in up to 43 languages. Through the extraction of meta-information in these articles, they provide

an aggregated view of the news, they allow to monitor trends and to navigate the news over time and even across languages. EMM-NewsExplorer additionally collects historical information about persons and organisations from the multilingual news, generates co-occurrence and quotation-based social networks, and more. All EMM applications were entirely developed at, and are being maintained by, the European Commission's Joint Research Centre (JRC) in Ispra, Italy.

The applications make combined use of a variety of text

analysis tools, including clustering, multi-label document classification, named entity recognition, name variant matching across languages and writing systems, topic detection and tracking, event scenario template filling, and more. Due to the high number of languages covered, linguistics-poor methods were used for the development of these text mining components. See the site <http://langtech.jrc.it/> for technical details and a list of publications.

The speaker will give an overview of the various applications and will then explain the workings of selected text analysis components.

Cyc Random Computer Thoughts:

<http://sw.opencyc.org/>

NURIKABE

<http://www.puzzle-nurikabe.com/>

Rules:

Find out whether each of the cells of the grid is "black" or "white".

The black cells form "the nurikabe": they must all be orthogonally contiguous (form a single polyomino), number-free, and contain no 2x2 or larger solid rectangles

The white cells form "islands": each number n must be part of an n-omino composed only of white cells. All white cells must belong to exactly one island; islands must have exactly one numbered cell.

Every resurrection destroys some dead person	A pardon is a type of forgiving	Most notably, rescuing is a type of helping	In a daily morning routine, the "going" step ends later than the "brushing of teeth" step
All ball bearing manufacturer makes ball bearings	The characteristic and frequent activity of a scientist is empirical testing	Engineering units are typically equipped with mine delivery systems	[dirty] is a larger value than [a little dirty]
Cleaning generally occurs in an emergency room	A rebound is a type of gaining possession in a game	Every daily routine effects can be felt for a few hours	No time lasts longer than forever

Random Computer Thoughts

Cyc is "world's largest and most complete general knowledge base and commonsense reasoning engine". Its open source version is called OpenCyc, having also a version for the Semantic Web. In order to generate Random

Computer Thoughts, Cyc randomly picks a concept inside its knowledge base, reasons about this concept and uses its English language capabilities to express the result of its reasoning process. The particular concept is followed

until Cyc runs out of ideas, after which it switches topic and the process continues. Here are some of Cyc's Random Computer Thoughts. To find out more concepts, follow the link on the left of the series of Cyc thoughts.

3				1	2	1	1		
1		3				2	2	2	
	4			4					
						4			2
2			3						
					3	1			
3			2						3
				2					
						5			
4									

Postojna Cave

The Postojna cave or "Postojnska jama" in Slovene, is a network of galleries, chambers and passages created by the Pivka River. The endemic olm, the largest troglodytic amphibian in the world, has his home there. You might not be able to see it though, because it prefers darkness (and tourists just love flashing it). The cave has

railway lines (since 1872), so one can admire the speleothems (calcite formations, stalactites and stalagmites), the cave "curtains" and also enjoy the train ride.

Trivia:

- It's the largest cave in the Classic Karst
- It's the most visited tourist cave in Europe: more than 31 million visitors in the last 188 years

- It's a 20,570 m long Karst cave
- The olm is a symbol of Slovenian natural heritage
- The castle built on top of the cave has a secret passage that leads into the cave
- Check out Postojna cave 360° at http://www.slovenia.info/?kraska_jama=450