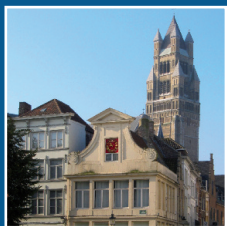## Hotel Martin's Brugge (www.martinshotels.com) - In Bruges, Belgium

By invitation only

Text mining STM Content:
New Perspectives, Emerging Solutions

Organized by the P-D-R Taskforce on Text mining

Core Programming Committee

From P-D-R: Henning P. Nielsen (Novo Nordisk), Andrew Clark (UCB), Martin Romacker (Roche)
From STM publishers: Carlo Scollo Lavizzari (STM), Duncan Campbell (Wiley), Jan bij de Weg (Elsevier)

Objectives of the meeting

Text mining and text mining tools have become an integrated part of standard information management in pharmaceutical companies, and the emergence of concepts such as "big data" and minable repositories or online data sets have accentuated the need for standards and business models in order to exploit the technological innovations taking place in this field.

P-D-R has for many years been partnering with the stakeholders in the information industry, and in particular with the STM publishers clarifying rights and licensing issues.

In 2011 the P-D-R Library & Copyright Topic Group arranged a meeting in Bruges, Belgium under the headline: **Text mining STM Content: Case Studies, Technical Challenges & Rights Issues**. meeting initiated the creation of two P-D-R taskforces I: *A task force to agree on a text for licensing of full text for text mining purposes.*
II: *A task force to identify and propose solutions for standardised technical solutions delivering XML data for text mining purposes.*
Taskforce I finalised its target in 2012 by joint publication of licensing clauses to be included in the P-D-R model license.
Taskforce II was joined with Taskforce I, and agreed, that in the present market place standardisation would be mainly market driven, so as a consequence the P-D-R members decided to work actively on the requirements from the industry side, for discussion during 2013.

The taskforce also agreed on the need for an update on the text mining marketplace since the last meeting in 2011, again bringing together:
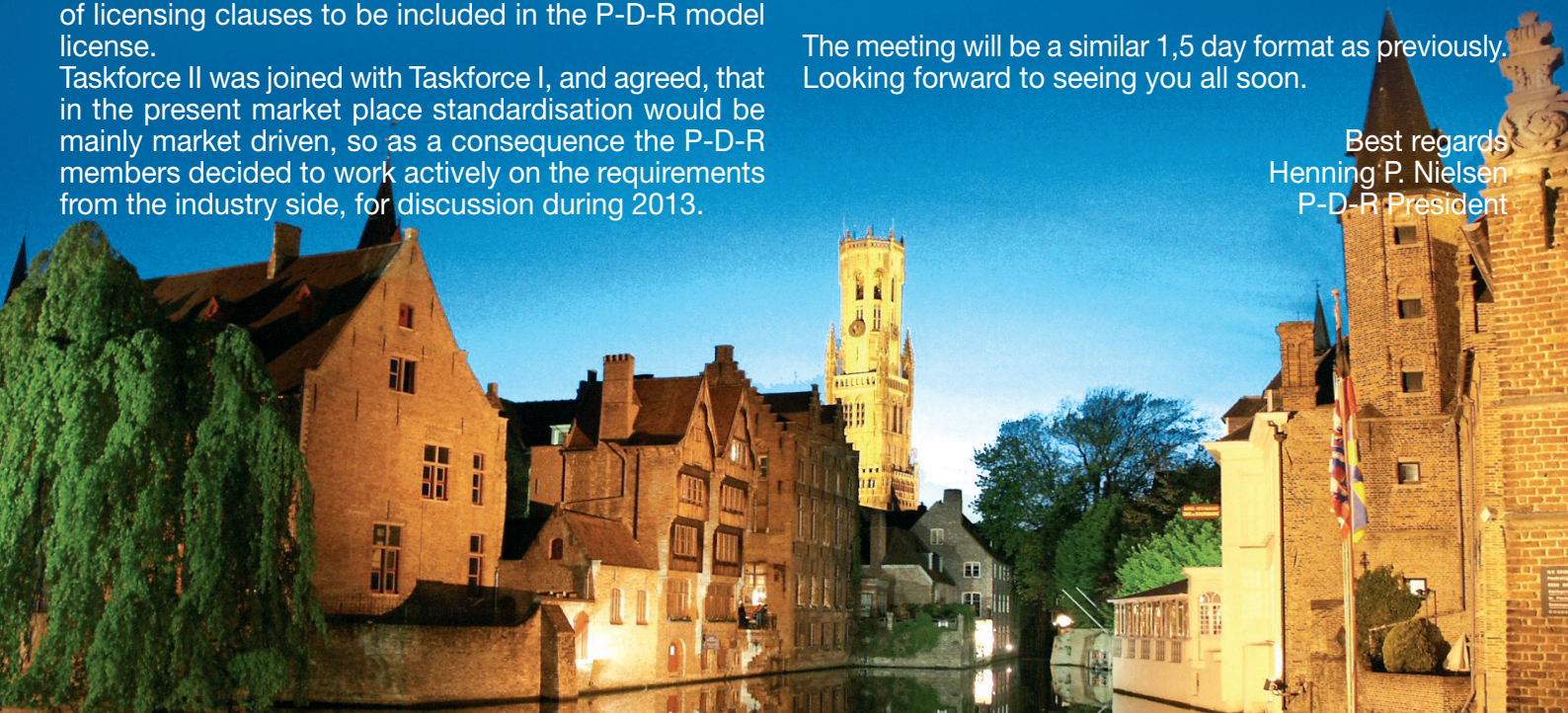
- information management executives and text analytic professionals from pharma companies;
- vendors of text analytics solutions;
- STM publishers;
- providers and/or aggregators of full text archives;
- information solution vendors;
- representatives from reproduction rights organizations (RROs);
- academic researchers working on text mining projects.

This should give us again the chance to meet, learn and understand the needs of all partners.

Based on such a common understanding of what corporate information centres and corporate text analytic facilities need to provide to their customers, knowing what solutions are currently in place, and identifying significant unmet needs, the Special Meeting is aimed to update all stakeholders on the text mining market place and facilitate discussions on innovations, standardisation needs and possible business models.

The meeting will be a similar 1,5 day format as previously. Looking forward to seeing you all soon.

Best regards
Henning P. Nielsen
P-D-R President

MEETING AGENDA

P·D·R
PHARMA DOCUMENTATION RING

Wednesday, November 27, 2013

19:30   **Reception at Hotel Martin's Brugge** - Bruges, Belgium

Thursday, November 28, 2013

Please note: Time slots include questions and answers

Morning session - Moderator: Henning P. Nielsen, Novo Nordisk

09:00   **Welcome** - *Henning P. Nielsen, Novo Nordisk, P-D-R President*

09:10   **Opening:** Text mining - the past, the present, the future - *Andrew Clark, UCB*

09:20   Text Mining - Vision and opportunities - *Key Note Speaker - Sophia Ananiadou, Manchester University*

10:00   Case Study from P-D-R 1: Utopia Documents@Roche: Text mining as enabler for next-generation PDF reading
Markus Bundschus, Roche

10:25   Case Study from P-D-R 2: Compliant, aggregated data for repurposing - optimal workflows and costs
Barbara Reissland, UCB

10:50   Case Study from P-D-R 3: Semantic technologies: Smart solutions in a world of information overflow
Torsten Zank, BASF

11:15   **Coffee Break**

11:35   Case Study from P-D-R 4: Full-text versus abstracts - Experiences from mining patents and literature
Wolfgang Thielemann, Bayer Health Care

12:00   Text mining open access and legal status / EU / UK
1) The end user perspective: needs and requirements - *Wilma Mossink, IFLA*
2) Legal status EU/UK and Licensing solutions - *Carlo Lavizzari, STM*

12:45   **Lunch**

Afternoon Session - Moderator: Duncan Campbell, Wiley

14:00   Text Mining Case Studies from Academia 1: Biomedical text mining at scale
Ulf Leser, Humboldt Universität zu Berlin

14:45   Text Mining Case Studies from Academia 2: Pragmatic text mining: From literature to electronic health records
Lars Juhl Jensen, NNF Centre for Protein Research, Copenhagen University

15:30   **Coffee Break**

16:00   Text Mining Case Studies from Publishers Perspective
Duncan Campbell, Wiley and Jan bij de Weg, Elsevier

16:20   Text Mining Case Studies: CCC Pilot Project
Kate Alzapiedi, RightsDirect and Babis Marmarnis, Copyright Clearance Center

16:40   Text Mining Case Studies: Cross Ref project - *Geoffrey Bilder, Cross Ref*

17:00   Software providers - Short statements - Panel - *Moderator: Carlo Lavizzari, STN*
- Averbis - *Philipp Daumke*
- IBM - *Stephen Boyer*
- Intomics - *Thomas Skøt Jensen*
- Linguamatics - *David Milward*
- Temis - *Stefan Geissler*
- Molecular Connections - *Jignesh Bhate*

18:00   **End of day 1 programme**

19:15   **HorseTram** (*departure 19:15 from Hotel Martin's Brugge*)

20:15   **Dinner at Restaurant Maximiliaan van Oostenrijk**

MEETING AGENDA

P·D·R
PHARMA DOCUMENTATION RING

Friday, November 29, 2013

Morning session - Moderator: Andrew Clark, UCB

08:30	**Wrap-up from Day 1** - *Andrew Clark, UCB*
08:40	Emerging tagging standards for text mining on full-text journal articles - P-D-R requirements *Martin Romacker, Roche*
09:00	Publishers' perspectives on technical issues and solutions - *Michiel van der Heyden, Springer*
09:20	What solutions are offered by aggregators - Panel - *Moderator: Martin Romacher, Roche* • Copyright Clearance Centre - *Babis Marmanis* • Infotrieve - *Mary Ging* • Quosa - *Andrew Carver* • Reprints Desk - *Peter Derycz* • EBSCO - *Hans-Peter Meulekamp* • ProQuest - *Henrik Bang* • Thomson Reuters - *Tim Miller*
10:10	**Coffee Break**
10:30	Talking about solutions and business models - *Moderator: Andrew Clark, UCB* *Break-out* Questions to groups based on previous discussions - formed by organising group before morning start and during coffee break - *Initiated/moderated by Andrew Clark, UCB*
12:00	Round-Table feedback and discussion by break-out group leaders based on discussions
12:45	Wrap-up & action items and Next steps
13:00	**Lunch**
14:00	**Departure**