# Advances in Automated Knowledge Base Construction

Fabian M. Suchanek
Max Planck Institute
for Informatics, Germany
suchanek@mpi-inf.mpg.de

James Fan
IBM Research
Almaden, CA, USA
fanj@us.ibm.com

Raphael Hoffmann
University of Washington
Seattle, WA, USA
raphaelh@cs.washington.edu

Sebastian Riedel
University College London
London, UK
sebastian.riedel@gmail.com

Partha P. Talukdar
Carnegie Mellon University
Pittsburgh, PA, USA
partha.talukdar@cs.cmu.edu

## ABSTRACT

Recent years have seen significant advances on the creation of large-scale knowledge bases (KBs). Extracting knowledge from Web pages, and integrating it into a coherent KB is a task that spans the areas of natural language processing, information extraction, information integration, databases, search and machine learning. Some of the latest developments in the field were presented at the AKBC-WEKEX workshop on knowledge extraction at the NAACL-HLC 2012 conference. This workshop included 23 accepted papers, and 11 keynotes by senior researchers. The workshop had speakers from all major search engine providers, government institutions, and the leading universities in the field. In this survey, we summarize the papers, the keynotes, and the discussions at this workshop.

## 1. INTRODUCTION

The advances in information extraction, machine learning, and natural language processing have led to the creation of large knowledge bases (KBs) from Web sources. Notable endeavors in this direction include Wikipedia-based approaches (such as YAGO [39], DBpedia [2], and Freebase), systems that extract from the entire Web (NELL [6], PROSPERA [27]) or from specific domains (Rexa [43]), and open information extraction approaches (TextRunner [4], PRISMATIC [46]). This trend has led to new applications that make use of semantics. Most prominently, all major search engine providers (Yahoo!, Microsoft Bing, and Google) nowadays experiment with semantic tools. The Semantic Web, too, benefits from the new approaches.

The AKBC-WEKEX workshop on Knowledge Extraction brought together leading experts from the field to discuss the latest developments in the area. Senior researchers from the search engine providers (Microsoft, Bing, and Yahoo!), as well as from the Defense Advance Research Project Agency of the United States (DARPA), and from leading universities gave invited talks on their newest research. In addition, the workshop invited regular paper submissions. The workshop focused exclusively on visionary submissions. The papers were only 4 pages long, and the emphasis of the workshop was on personal interaction and visionary ideas rather than on technical details and experiments. In total, the workshop attracted 39 submissions, of which we accepted 23. All papers were reviewed by 3 reviewers. Our program committee comprised 24 leading researchers in the area of knowledge extraction. Three submissions (the "Top-3") were deemed particularly creative, and given oral presentations. All other papers were presented in a spotlight talk, and through posters. One submission was awarded the Best Paper Award.

In this paper, we survey all keynotes and all vision papers, in order to give an overview of some of the new developments in the field of knowledge base construction.

## 2. KEYNOTE TALKS

**Professor Tom Mitchell** is the chair of the Machine Learning Department at Carnegie Mellon University (CMU). He presented the Never-Ending Language Learning system NELL [6]. NELL reads natural language text on the Web and feeds it into a knowledge base. NELL has been running continuously since January 2010, and has already accumulated 15 million candidate beliefs. The central technique of NELL is coupled learning: multiple learners constrain each other, and thus continuously improve their joint performance. NELL has learners for textual patterns, argument types, morphology, and Horn rules. NELL can also learn new relation types by clustering pairs of instances that appear together in textual patterns. A recent addition to

NELL is an inference technique based on random walks [19]: The system can find out that certain sequences of relation labels on a path in the knowledge base always lead to the same relation label. This allows it to suggest that other, similar, paths in the knowledge base should also lead to that label. Thereby, NELL can auto-complete its knowledge.

**Professor Steffen Staab** is the director of the Institute for Computer Science at Koblenz University in Germany. In his talk, he first pointed out the new paradigms of knowledge representation on the Semantic Web (SW): First, there is no central authority on the SW. Therefore, knowledge on the SW is inherently modular. Second, the data of the SW is distributed across different data providers, and no longer defined in one place. Third, all data on the SW is interconnected. It is possible to walk from one data item by any provider to any other data item by any other provider. Fourth, knowledge on the SW is continuously growing and, by design, extensible. It is never claimed to be complete. Fifth, data on the SW is large, but computationally lightweight. Last, the SW has a democratic mechanism for its choice of vocabulary, by which every vocabulary is equally valid, but popular vocabulary is more useful (and thus becomes more popular). Professor Staab summarizes these tenets as the "Lego style" of knowledge representation. He went on to present several examples that apply these paradigms to knowledge extraction, among others a visualization of urban parameters on Open Street Maps, and faceted browsing with linked data. His newest work is on SXPath [11], a spatial extension of XPath, which allows querying an HTML page also for spatial relationship of components.

**Dr. Bonnie Dorr**, program manager from DARPA, gave an overview of the language research programs at DARPA. There are four projects that encompass five areas of language processing. MADCAT focuses on OCR, RATS deals with speech recognition, BOLT covers speech recognition, translation and retrieval, and the new DEFT project will deal with retrieval as well as deep language understanding. Since the first three projects are already underway, Dr. Dorr focused on the new DEFT project. The goal of DEFT is to "see through language to meaning in text" so that the unstated semantics of sentences can be inferred. This requires a variety of natural language technology, such as deep semantic understanding of people, events, relations, etc. DEFT can be used for applications such as smart filtering, relational analysis and anomaly analysis. More detailed information on DEFT can be found at `http://www.darpa.mil/Our_Work/I2O/Programs/`.

**Dr. Nilesh Dalvi** from Yahoo Research presented an overview of his work on Domain-Centric Information Extraction [10][9]. While much research is concerned with scaling extraction to different domains, Dalvi's work focuses on scaling extraction to a very large number of sources while keeping domains fixed. In the first part of his talk, Dalvi showed results of a large-scale analysis of data on the Web that led to two key insights: First, to build a reasonably complete database for a domain one needs to got to the long tail of websites, even for domains where there are well-established aggregator sites. Second, the information within a domain tends to be well-connected on the Web, with a high degree of redundancy and overlap. In the second part of his talk, Dalvi presented a system that populates a given schema from the entire Web based on wrapper induction. The system automatically identifies and clusters pages, induces a very large number of wrappers, and uses a probabilistic model to rank wrappers and extractions. On several domains, the extractions reached an almost perfect F1 score.

**Professor Oren Etzioni** is the director of the Turing Center at the University of Washington. He presented an overview of his work on Open Information Extraction (OpenIE) [13][4][24]. OpenIE is motivated by the goal to perform machine reading at Web scale, and assumes that only an unsupervised approach that is not limited by a pre-determined ontology can succeed. Although OpenIE systems target only an easy-to-understand subset of English language, they scale linearly and extract orders of magnitude more relations than other extraction systems. The output of OpenIE can be linked to an ontology when required [22]. Etzioni further defined three areas in which OpenIE is making an impact: applications of next-generation search that are based on OpenIE [12], a series of public textual textual resources based on OpenIE [3], and work that demonstrates that it is possible to do reasoning over OpenIE extractions [37]. The latter includes work on learning synonyms, inference rules, and argument types. For future work, he sees an opportunity in connecting OpenIE with Linked Open Data (LOD).

**Dr. James Fan**, research staff member from IBM Research and one of the main contributors to the Watson question answering system presented an overview of the DeepQA project for the Jeopardy! Challenge. After a brief description of the challenge of answering Jeopardy! questions and the overall DeepQA architecture, he delved into details of the hypothesis scoring aspect of the Watson system and the parts that are highly relevant to the workshop: relation extraction and structured knowledge utilization. Watson uses a statistical relation extraction method that is trained on DBpedia data. It uses Wikipedia-based structured sources, such as YAGO [39] and DBpedia [2], as well as an automatically built shallow proposition repository, PRISMATIC [46]. The knowledge from the structured sources is used for type coercion, answer merging, spatial reasoning and evidence diffusion. Details of the Watson system can be found in [45].

**Dr. Patrick Pantel** from Microsoft Research gave an overview of his work on mining action intents in Web search [23][32]. For a large fraction of Web search queries, a user seeks to find information about an entity or transact on the entity (e.g. buy); by understanding the underlying query intent one can provide a rich search experience. Pantel showed that the entities referenced in queries belong to only a small set of entity types, and that the user is typically interested in only a small set of actions tied to these entity types. Pantel therefore tried to define a theory for how actionable queries are generated. He evaluated a series of generative probabilistic models to predict a set of actions from a query, taking into account information such as clicked hosts or types of entities contained in the query. Several models yielded large improvements over a baseline and can be used to reliably generate reasonable actions for queries.

**Professor Andrew McCallum** from UMass Amherst presented an overview of his work on probabilistic databases for knowledge base construction. His talk started with the motivating example of populating a database of all scientists in the world based on papers, patents, Web pages etc. Professor McCallum then contrasted the traditional approach of populating the database with results of an IE pipeline (with its cascading errors) to his group's approach of performing IE inference "inside the database". In this approach the database continuously infers new facts and improves its beliefs for old ones based on incoming evidence. He coined this an Epistemological Database, indicating that the database doesn't observe or store the truth about entities and relations; it must infer the truth from available evidence. He showed how this view can be extended to (possibly erroneous) human edits [42]; a user would not directly edit a database; instead, she provides additional mentions of a fact, and these mentions are used alongside textual mentions when inferring truth. The remainder of his talk discussed how inference and learning can scale in such a scenario [43], and an approach to relation extraction with universal schema based on matrix factorization [44]. He stressed that all of the above are implemented on top of UMass' probabilistic programming framework FACTORIE, a Scala library for factor graphs [26].

**Professor Eduard Hovy** from the Information Science Institute at the University of Southern California (currently at Carnegie Mellon University) argued that even after research over the years in the area of automatic knowledge harvesting from text, we have very limited understanding of some of the fundamental questions in this area: how many concepts and relations are there in a given domain? what are those relations? how do we evaluate the extracted knowledge? and how do we store and integrate knowledge harvested by different teams and make them publicly available and useful? Prof. Hovy also provided some initial attempts at answering these question based on his prior research [40][21]. In discussions following the talk, one of the questions raised was whether we really need to answer these questions in their entirety before something useful could be done with current state-of-the-art (e.g., search engines have started to surface results based on automatically harvested knowledge, IBM's Watson system, which went on to defeat humans in the game of Jeopardy, benefited from having access to automatically harvested knowledge bases such as YAGO [39]).

**Dr. Fernando Pereira**, Director of Research at Google Inc., outlined challenges and opportunities for automatic knowledge harvesting in the context of web search. He observed that the Web search engines have started a recent shift to think in terms of concepts and not just terms (i.e., "Things, not Strings", as he pointed out). Google's Knowledge Graph is one instance of this phenomenon. In order to fully realize this goal, Dr. Pereira mentioned the need for effective and scalable solutions to many of the problems which are of central importance to the knowledge harvesting community (e.g., mapping terms into concepts, identifying relationship among concepts, etc.). He also emphasized on the need for linguistic analysis (e.g., coreference resolution, dependency parsing, etc.) for these problems. Finally, he presented learning and inference over graphs as a unifying theme to address many of these problems, with supporting evidence from recent research carried out by him and his co-authors [19].

**Professor Christopher Re**, Assistant Professor at the University of Wisconsin Madison, presented Hazy, his project on integrating statistical processing techniques with data processing systems. At the heart of his work is the hypothesis that breakthroughs in data analysis will not necessarily stem from new algorithms, but from new ways of combining, deploying and maintaining algorithms. Professor Re showed three demos developed with this hypothesis in mind. The first was DeepDive [31], a web-scale KB population engine that crawls 500M webpages, 400k videos and 20k books. At the heart of this system is inference on a large-scale factor graph defined by Markov Logic Networks. The second demo was GeoDeepDive, a system that populates a domain specific knowledge base from geo-scientific literature. His final system, Ancient-Texts, supports humanities researchers by populating a KB from historic texts. As OCR is a core part of this endeavour, Professor Re also presented Staccato [18], a method his group developed for maintaining the uncertainty provided by the OCR engine while maintaining a minimal memory footprint.

**Summary.** Our keynote speakers presented several large-scale endeavors on automated knowledge base construction. Particular emphasis was put on probabilistic reasoning and the scalability of the approaches. At the same time, the keynote speakers also provided application scenarios, ranging from enhanced keyword search to faceted search and data visualization. Thus, it seems that the time has come where both the supporting technology and the application technologies are mature enough to allow for the rise and use of automatically constructed knowledge bases.

## 3. VISION PAPERS

**Natural language.** Language is the basis of the vast majority of documents on the Web. Therefore, it is important to understand its basic properties before venturing into fact extraction. We had 3 papers that took a closer look at natural language. [29] provides a complete syntactic annotation of the Gigaword corpus, thereby creating a standard annotated resource for other researchers to work on. [20] analyzed how different cultures (English-speaking and Mandarin-speaking) use figurative language. In English, e.g., a pig is the archetype of something lazy, while in Chinese, it is the archetype of something happy (which is not contradictory). Finally, [15] derives the frequency of habitual human activities from phrases on the Web. For example, people sleep every day, but take a taxi only a few times a year.

**Disambiguation.** Mapping mentions in documents to entities in a knowledge base is a fundamental task for automatic web scale knowledge base construction. There are four papers presented that focus on this topic. [14] proposes a generative model that learns distributional semantics by performing entity linking, and the resulting distributional semantics is added to knowledge base entities. [38] utilizes discourse context to help disambiguate mentions. [22] investigates entity linking over millions of extractions, and uses corpus level features, such as collective context, to help disambiguate mentions. [36] presents a use-case of entity linking in news articles, where the set of relevant entities changes over time.

**Unsupervised learning.** Approaches to automatic knowledge base construction often require a prohibitive amount of human input. Therefore, much research is focused on unsupervised techniques. Five papers at our workshop fall in this category. [8], a top 3 paper, proposes a new low-dimensional representation for entities occurring on the Web, and shows that it can benefit several tasks such as set instance acquisition. [34] proposes an algorithm that is able to learn the main concepts of event summaries in a domain. [16] reflects on previous work on grammar discovery and discusses implications for future work on unsupervised information extraction. [7] summarizes ongoing work on a question answering system for biology questions, which is based on a KB of semi-structured assertions. Finally, [1] proposes an open information extraction system that can extract n-ary relations. Using the output of a typed dependency extractor, its extractions are more precise than those of a state-of-the-art open information extraction system for binary relations.

**Probabilistic approaches.** Knowledge extraction works in a context of noise, ambiguity and uncertainty. Probabilistic approaches are therefore essential for KB construction, and with seven papers this topic has the largest number of papers in this workshop. The authors of [30] present their vision of a distributed inference algorithm based on conict graph construction and hypergraph sampling. [17] shows how Tractable Markov Logic, a subset of Markov Logic that lends itself to more efficient inference, can be used for KB construction. In [3] the authors adapt the notion of n-grams from language models to models that predict relations based on other relations in the same context. [41] shows a vision of an automatic knowledge base construction system that uses statistical inference to integrate extraction, reasoning, and human feedback. A core technique in their system is MCMC, a sampling technique that scales with the number of factors a proposal involves. [35] improve the efficiency of this method by sub-sampling the set of involved factors. [44] propose Universal Schemas that incorporate both structured relations and OpenIE surface patterns, and argue for collaborative filtering methods to probabilistically missing facts. [42], a top 3 paper, describes a new framework for integrating human edits and IE. It is based on the principle that users should not directly alter truth, but provide their input as mentions that are used to infer truth in the probabilistic database.

**Time.** In many cases, time plays a role for the correctness of a fact. For example, the fact *presidentOf(Bill Clinton, USA)* was true during the period 1993-2001, but is no longer true today. This problem of extending a KB along the temporal dimension has been called Temporal Slot Filling (TSF) or Temporal Scoping. A method for improved TSF was presented in [33]. The key contributions were to improve the quality of distantly supervised training data, and use a combination of feature selection and self-training to expand a small set of initial labeled instances. A related problem of fact extraction in real-time was explored in [28], one of the top three papers at the workshop. In contrast to traditional knowledge harvesting techniques, which tend to be batch-oriented and thereby not quite suitable for rapidly changing data (e.g., twitter streams), [28] outlined the challenges and opportunities for the KB construction community in harvesting knowledge from frequently changing data. This contribution earned the paper the *Best Paper Award* of the workshop. A sketch of a prototype system to address these challenges was also presented in [36].

**Evaluation.** Once a KB has been constructed, it is necessary to evaluate its quality. Our workshop had two papers on this topic. [5] focuses on the automatic evaluation of relation extraction using public online database and large web corpus. [25] evaluates the quality of knowledge bases through a set of sample queries.

## 4. CONCLUSION

The AKBC-WEKEX workshop on knowledge extraction has brought together researchers from universities, industry, and government institutions. Together, we have pointed out new trends in the field of knowledge base construction. Our keynote speakers showed that large-scale endeavors on automated knowledge base construction are successful both in academia and in the search engine business. Our vision papers have pointed out new research challenges at the nexus of natural language documents and structured knowledge bases. Particular attention has been paid to probabilistic knowledge extraction and to the evolution of knowledge over time. The large variety of ideas brought forward by the vision papers shows us that many exciting research questions are still to be answered for the automated construction of knowledge bases.

## 5. REFERENCES

[1] Alan Akbik And Alexander Löser: Kraken: N-ary Facts In Open Information Extraction. In AKBC-WEKEX 2012.

[2] Sören Auer, Jens Lehmann: What have Innsbruck and Leipzig in common? Extracting Semantics from Wiki Content. In ESWC 2007.

[3] Niranjan Balasubramanian, Stephen Soderland, Mausam And Oren Etzioni: Rel-grams: A Probabilistic Model Of Relations In Text. In AKBC-WEKEX 2012.

[4] Michele Banko, Michael J. Cafarella, Stephen Soderland, Matt Broadhead And Oren Etzioni: Open Information Extraction from the Web. In IJCAI 2007.

[5] Mirko Bronzi, Zhaochen Guo, Filipe Mesquita, Denilson Barbosa And Paolo Merialdo: Automatic Evaluation Of Relation Extraction Systems On Large-scale. In AKBC-WEKEX 2012.

[6] A. Carlson, J. Betteridge, R.C. Wang, E.R. Hruschka Jr., and T.M. Mitchell: Coupled Semi-Supervised Learning for Information Extraction. In WSDM 2010.

[7] Peter Clark, Phil Harrison, Niranjan Balasubramanian, and Oren Etzioni: Constructing A Textual Kb From A Biology Textbook. In AKBC-WEKEX 2012.

[8] Bhavana Dalvi, William Cohen, and Jamie Callan: Collectively Representing Semi-structured Data From The Web. In AKBC-WEKEX 2012.

[9] Nilesh N. Dalvi, Ravi Kumar And Mohamed A. Soliman: Automatic Wrappers for Large Scale Web Extraction. In PVLDB, Volume 4, Number 4, 2011.

[10] Nilesh N. Dalvi, Ashwin Machanavajjhala And Bo Pang: An Analysis of Structured Data on the Web. In PVLDB, Volume 5, Number 7, 2012.

[11] Linda d'Oro, Massimo Ruffolo, Steffen Staab: SXPath - Extending XPath towards Spatial Querying on Web Documents. In PVLDB 2010.

[12] Oren Etzioni: Search needs a shake-up. In Nature, 476: 25-26, August 4, 2011.

[13] Oren Etzioni, Anthony Fader, Janara Christensen, Stephen Soderland, and

Mausam: Open Information Extraction: the Second Generation. In IJCAI 2011.

[14] Matt Gardner: Adding Distributional Semantics To Knowledge Base Entities Through Web-scale Entity Linking. In AKBC-WEKEX 2012.

[15] Jonathan Gordon And Lenhart Schubert: Using Textual Patterns To Learn Expected Event Frequencies. In AKBC-WEKEX 2012.

[16] Ralph Grishman: Structural Linguistics And Unsupervised Information Extraction. In AKBC-WEKEX 2012.

[17] Chlo Kiddon and Pedro Domingos: Knowledge Extraction And Joint Inference Using Tractable Markov Logic. In AKBC-WEKEX 2012.

[18] Arun Kumar and Christopher Re: Probabilistic Management of OCR using an RDBMS. In PVLDB 2012.

[19] N. Lao, T.M. Mitchell, W.W. Cohen: Random Walk Inference and Learning in A Large Scale Knowledge Base. In EMNLP 2011.

[20] Bin Li, Jiajun Chen And Yingjie Zhang: Web Based Collection And Comparison Of Cognitive Properties In English And Chinese. In AKBC-WEKEX 2012.

[21] Zornitsa Kozareva, Ellen Riloff and Eduard Hovy. Semantic Class Learning from the Web with Hyponym Pattern Linkage Graphs. In ACL-HTL 2008.

[22] Thomas Lin, Mausam And Oren Etzioni: Entity Linking At Web Scale. In AKBC-WEKEX 2012.

[23] Thomas Lin, Patrick Pantel, Michael Gamon, Anitha Kannan, and Ariel Fuxman: Active Objects: Actions for Entity-Centric Search. In WWW 2012.

[24] Mausam, Michael Schmitz, Robert Bart, Stephen Soderland, and Oren Etzioni: Open Language Learning for Information Extraction. In EMNLP 2012.

[25] James Mayfield And Tim Finin: Evaluating The Quality Of A Knowledge Base Populated From Text. In AKBC-WEKEX 2012.

[26] Andrew McCallum, Karl Schultz, Sameer Singh. FACTORIE: Probabilistic Programming via Imperatively Defined Factor Graphs. In NIPS 2009.

[27] Ndapandula Nakashole, Martin Theobald and Gerhard Weikum: Scalable Knowledge Harvesting with High Precision and High Recall. In WSDM 2011.

[28] Ndapandula Nakashole And Gerhard Weikum: Real-time Population Of Knowledge Bases: Opportunities And Challenges. In AKBC-WEKEX 2012.

[29] Courtney Napoles, Matthew Gormley And Benjamin Van Durme: Annotated Gigaword. In AKBC-WEKEX 2012.

[30] Mathias Niepert, Christian Meilicke And Heiner Stuckenschmidt: Towards Distributed Mcmc Inference In Probabilistic Knowledge Bases. In AKBC-WEKEX 2012.

[31] Feng Niu, Ce Zhang, Christopher Re, and Jude Shavlik, DeepDive: Web-scale Knowledge-base Construction using Statistical Learning and Inference VLDS, 2012.

[32] Patrick Pantel, Thomas Lin, Michael Gamon: Mining Entity Types from Query Logs via User Intent. In ACL 2012.

[33] Suzanne Tamang And Heng Ji: Relabeling Distantly Supervised Training Data For Temporal Knowledge Base Population. In AKBC-WEKEX 2012.

[34] Horacio Saggion: Unsupervised Content Discovery From Concise Summaries. In AKBC-WEKEX 2012.

[35] Sameer Singh, Michael Wick And Andrew Mccallum: Monte Carlo Mcmc: Efficient Inference By Sampling Factors. In AKBC-WEKEX 2012.

[36] Rosa Stern And Benoit Sagot: Population Of A Knowledge Base For News Metadata From Unstructured Text And Web Data. In AKBC-WEKEX 2012.

[37] Stefan Schoenmackers, Oren Etzioni, and Daniel S. Weld: Scaling Textual Inference to the Web. In EMNLP 2008.

[38] Veselin Stoyanov, James Mayfield, Tan Xu, Douglas Oard, Dawn Lawrie, Tim Oates And Tim Finin: A Context-aware Approach To Entity Linking. In AKBC-WEKEX 2012.

[39] Fabian M. Suchanek, Gjergji Kasneci and Gerhard Weikum: YAGO - A Core of Semantic Knowledge. In WWW 2007.

[40] S. Tratz and E.H. Hovy. 2010. A Taxonomy, Dataset, and Classifier for Automatic Noun Compound Interpretation. In ACL 2010.

[41] Daisy Zhe Wang, Yang Chen, Sean Goldberg, Christan Grant And Kun Li: Automatic Knowledge Base Construction Using Probabilistic Extraction, Deductive Reasoning, And Human Feedback. In AKBC-WEKEX 2012.

[42] Michael Wick, Karl Schultz And Andrew Mccallum: Human-machine Cooperation: Supporting User Corrections To Automatically Constructed Kbs. In

AKBC-WEKEX 2012.

[43] Michael Wick, Sameer Singh, Andrew McCallum. A Discriminative Hierarchical Model for Fast Coreference at Large Scale. In ACL 2012.

[44] Limin Yao, Sebastian Riedel And Andrew McCallum: Probabilistic Databases Of Universal Schema. In AKBC-WEKEX 2012.

[45] Watson. In IBM Journal of Research and Development, Volume 56, Issue 3:4, May-June 2012.

[46] J. Fan, A. Kalyanpur, D.C. Gondek,D.A. Ferrucci: Automatic knowledge extraction from documents. In IBM Journal of Research and Development, Volume 56, Issue 3:4, May-June 2012