AMIA 2011
SHARPn Colleague Participation

Sessions

Saturday - Oct 22, 2011

8:30 AM-4:30 PM
Monroe (Washington Hilton)
T07: Clinical Classifications and Biomedical Ontologies: Terminology Evolution, Principles, and Practicalities
Christopher G. Ghute, Mayo Clinic, James J. Cimino, National Library of Medicine, and Mark Musen, Stanford University

8:30 AM-12:00 PM
Jefferson East (Washington Hilton)
T03: Practical Modeling Issues Representing Coded and Structured Patient Data in EHR Systems
Stanley Huff, Intermountain Healthcare

8:30 AM-12:00 PM
Lincoln East (Washington Hilton)
T06: An Introduction to Clinical Natural Language Processing Part 1
Sponsor: NLP WG

1:00 PM-4:30 PM
Jefferson East (Washington Hilton)
T13: An Introduction to Clinical Natural Language Processing, Part 2: Interactive demonstrations of fundamental concepts with available open source tools
Sponsor: NLP WG

Sunday - Oct 23, 2011

8:30 AM-12:00 PM
Georgetown West (Washington Hilton)
T15: Embracing Healthcare IT Standards in the World of Meaningful Use
Charles Jaffe, HL7, Rebecca Kush, CDISC, Dixie Baker, SAIC, Blackford Middleton, Harvard-Partners; Chris Chute, Mayo Clinic, Stanley Huff, Intermountan Healthcare, and Robert Dolin, Semantically Yours, LLC

Papers & Presentations

Sunday, Oct 23

3:30PM

Session Title: S04 - Panel: Shared Annotated Resources for the Clinical Domain
Session Type: Panel
Location: Jefferson West (Washington Hilton

AMIA-0537-A2011. Shared Annotated Resources for the Clinical Domain

G. Savova[1]; W. Chapman[2]; N. Elhadad[3]; M. Palmer[4]
1. Harvard University, Boston, MA, United States.
2. Division of Biomedical Informatics, University of California San Diego, La Jolla, CA, United States.
3. Biomedical Informatics, Columbia University, New York, NY, United States.
4. Linguistics, University of Colorado at Boulder, Boulder, CO, United States.

Natural Language Processing (NLP) of the clinical narrative has been a major effort within medical informatics. Advances, however, have been hampered by the lack of shared, large annotated corpora to be leveraged for methods development and system evaluations. In the general domain, the gold-standard annotated Penn Treebank (PTB) fostered truly revolutionary advances. Within the clinical domain, there are several recent, complementary initiatives to create shared annotated resources – the Shared Annotated Clinical Resource project, Strategic Health Advanced Research Project Area 4, Milti-source Integrated Platform for Answering Clinical Questions, the Temporal Relations in the Clinical Domain, Ontology Development and Information Extraction, and the Integrating Informatics and Biology to the Bedside initiatives. In each, care was taken to ensure common annotation schemas and guidelines, all compatible with PTB. The combined resources will contain 1.5 million tokens and will be available to the research community. Their availability is expected to energize the clinical NLP community as well as involve the general NLP community into porting best methods and practices to healthcare. The panel will report on progress on these efforts, discuss the use of existing community-adopted conventions and domain-specific types of annotations, as well as the implications of making the annotated corpus publicly available.

Monday, Oct 24

1:45PM

Session Title: S28 - Papers: Terminologies & Ontologies
Session Type: Paper
Location: Gunston (Washington Hilton)
AMIA-0482-A2011. An OWL Meta-Ontology for Representing the Clinical Element Model

C. Tao[1]; C. Parker; T. Oniki; J. Pathak[2]; S. Huff[3]; C. Chute[4]
1. Health Sciences Research, Mayo Clinic, Rochester, MN, United States.
2. Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, United States.
3. Informatics, Intermountain Healthcare, Murray, UT, United States.
4. Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, United States.

The Clinical Element Model (CEM) is a strategy designed to represent logical models for clinical data elements to ensure unambiguous data representation, interpretation, and exchange within and across heterogeneous sources and applications. The current representations of CEMs have limitations on expressing semantics and formal definitions of the structure and the semantics. Here we introduce our initial efforts on representing the CEM in OWL, so that the enrichment with OWL semantics and further semantic processing can be achieved in CEM. The focus of this paper is the CEM meta-ontology where the basic structures, the properties and their relationships, and the constraints are defined. These OWL representation specifications have been reviewed by CEM experts to ensure they capture the intended meaning of the model faithfully.

3:30 PM

Session Title: S37 - Papers: Information Retrieval & Text Mining
Session Type: Paper
Location: Lincoln West (Washington Hilton)

AMIA-0631-A2011. The MiPACQ Clinical Question Answering System

B. Cairns[1]; R. Nielsen[2]; J. Masanz[3]; J. Martin[4]; M. Palmer[5]; W. Ward[6]; G. Savova[7]
1. University of Colorado at Boulder, Boulder, CO, United States.
2. Institute of Cognitive Science, University of Colorado at Boulder, Boulder, CO, United States.
3. Mayo Clinic College of Medicine, Rochester, MN, United States.
4. Computer Science, University of Colorado at Boulder, Boulder, CO, United States.
5. Linguistics, University of Colorado at Boulder, Boulder, CO, United States.
6. Institute of Cognitive Science, Center for Computational Language and Education Research, University of Colorado at Boulder, Boulder, CO, United States.
7. Children's Hospital Boston and Harvard Medical School, Boston, MA, United States.

The Multi-source Integrated Platform for Answering Clinical Questions (MiPACQ) is a multi-corpus QA pipeline that integrates a variety of information retrieval and natural language processing systems into an extensible question answering system. We present the system's architecture and an evaluation of MiPACQ on a human-annotated evaluation dataset based on the Medpedia health and medical encyclopedia. Compared with our baseline information retrieval system, the MiPACQ rule based system demonstrates 84% improvement in Precision at One and the MiPACQ machine-learning-based system demonstrates 134% improvement. Other performance metrics including mean reciprocal rank and area under the precision/recall curves also showed significant improvement, validating the effectiveness of the MiPACQ design and implementation.

5:15 PM

Session Title: Poster Session 1
Session Type: Poster
Location: Columbia Hall
AMIA-0817-A2011. Normalized Representation of Data Elements for Phenotype Cohort Identification in the Electronic Health Record

C. Tao[1]; S. Welch[2]; W. Wei[3]; T. Oniki; C. Parker; J. Pathak[4]; S. Huff[5]; C. Chute[6]
1. Health Sciences Research, Mayo Clinic, Rochester, MN, United States.
2. Biomedical Informatics, University of Utah, Sandy, UT, United States.
3. Institute of Health Informatics, University of Minnesota Twin Cities & Mayo Clinic, Twin Cities, MN, United States.
4. Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, United States.
5. Informatics, Intermountain Healthcare, Murray, UT, United States.
6. Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, United States.

This paper introduces our efforts on representing data elements in phenotyping algorithms using the Clinical Element Model (CEMs). The data elements in phenotyping algorithms and the CEMs were designed for different purposes so they are represented differently in some cases. We have identified gaps for consideration in the design of our data normalization and high throughput phenotyping components within SHARP secondary use of EHR

Session Title: Poster Session 1
Session Type: Poster
Location: Columbia Hall
AMIA-0705-A2011. RPE: A Process Approach to Linking Researchers, Providers and Patients

M. Arratoon[1]; L. Bain[2]; R. Kush[2]; J. Aerts[3]
1. GE Healthcare, Burlington, VT, United States.
2. CDISC, Winston Salem, NC, United States.
3. XML4Pharma, , Germany.

CDISC's BRIDG and CDASH specifications and ONC's SHARP Program are efforts to standardize clinical research data and the secondary use of EHR data. But data specifications alone cannot capture the site-based processes that define the study itself: activities such as patient enrollment, visit scheduling, and ad hoc data capture. CDISC and IHE have developed an integration standard (or profile) called Retrieve Process for Execution (RPE) which enables a collaborative automation of these protocol-specified activities.

10:30AM

Session Title: S49 - Panel: Advances in Clinical Question Answering: Watson meets Healthcare
Session Type: Panel
Location: International Ballroom Center (Washington Hilton)

AMIA-1008-A2011. Advances in Clinical Question Answering: Watson meets Healthcare

J. Hurdle[1]; G. Savova[2]; M. Kohn[3]; J. Cimino[4]; A. Cohen[5]; D. Demner-Fushman[6]; R. Nielsen[7]; J. Patrick[8]; H. Yu[9]
1. Biomedical Informatics, University of Utah, Salt Lake City, UT, United States.
2. Harvard University, Boston, MA, United States.
3. Research, IBM, New York, NY, United States.
4. laboratory for Informatics Development, National Institutes of Health - Clinical Center, Bethesda, MD, United States.
5. Department of Medical Informatics & Clinical Epidemiology, Oregon Health & Science University, Portland, OR, United States.
6. LHNCBC, US National Library of Medicine, Bethesda, MD, United States.
7. Institute of Cognitive Science, University of Colorado at Boulder, Boulder, CO, United States.
8. School of Information Technologies, University of sydney, Sydney, Australia.
9. University of Wisconsin Milwuakee, Milwaukee, WI, United States.

This panel offers a thought-provoking discussion on a topic that stirred up an unprecedented amount of chatter in the AMIA Natural Language Processing Working Group community. Our goal is to review the state-of-the-art of clinical question answering systems (cQA) in light of the brilliantly showcased achievement of IBM's Watson QA system on the Jeopardy! game show. We take a nontraditional approach to this panel's format by briefly describing three important cQA tools, followed by three perspectives on the clinician's view of cQA in the world of life-after-Watson. Setting the stage, a lead IBM scientist will provide an overview of IBM's Watson technology and their healthcare plans. The panel will leave 20 minutes for a lively discussion with the audience. The learning objectives for the panel: 1) Attendees will be able to describe the special nature of clinical question asking at the point of care; 2) Attendees will understand how to compare and contrast three state-of-the-art clinical question answering systems; 3) Moving beyond the recent publicity, attendees will leave with an understanding of how IBM's technology is being re-purposed for healthcare. This panel is sponsored by the AMIA NLP WG.

1:45PM

Session Title: S65 - Papers: Patient Safety, Medication Lists and Terminologies
Session Type: Paper
Location: Gunston (Washington Hilton)
AMIA-0709-A2011. Using RxNorm and NDF-RT to Classify Medication Data Extracted from Electronic Health Records: Experiences from the Rochester Epidemiology Project

J. Pathak[1]; S. Murphy[1]; B. Wilaert[1]; H. Kremers[2]; B. Yawn[3]; W. Rocca[2]; C. Chute[4]
1. Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, United States.
2. Epidemiology, Mayo Clinic, Rochester, MN, United States.
3. Olmsted Medical Center, Rochester, MN, United States.
4. Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, United States.

RxNorm and NDF-RT published by the National Library of Medicine (NLM) and Veteran's Affairs (VA), respectively, are two publicly available Federal Medication drug terminologies. In this study, we evaluate the applicability of RxNorm and NDF-RT for extraction and classification of medication data retrieved using structured querying and natural language processing techniques from electronic health records at two different medical centers within the Rochester Epidemiology Project (REP). Specifically, we explore how mappings between RxNorm concept codes and NDF-RT drug classes can be leveraged for hierarchical organization and grouping of REP medication data, identify gaps and coverage issues, and analyze the recently released NLM's NDF-RT Web service API. Our study concludes that RxNorm and NDF-RT can be applied together for classification of medication extracted from multiple EHR systems, although several issues and challenges remain to be addressed. We further conclude that the Web service API's developed by the NLM provide useful functionalities for such activities.

3:30 PM

Session Title: S72 - Papers: CRI: Aggregation & Clustering
Session Type: Paper
Location: International Ballroom West (Washington Hilton)
AMIA-0910-A2011. The SHARPn Project on Secondary Use of Electronic Medical Record Data: Progress, Plans, and Possibilities

C. Chute[1]; J. Pathak[2]; G. Savova[3]; K. Bailey; M. Schor; L. Hart; C. Beebe; S. Huff[4]
1. Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, United States.
2. Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, United States.
3. Harvard University, Boston, MA, United States.
4. Informatics, Intermountain Healthcare, Murray, UT, United States.

SHARPn is a collaboration among 16 academic and industry partners committed to the production and distribution of high-quality software artifacts that support the secondary use of EMR data. Areas of emphasis are data normalization, natural language processing, high-throughput phenotyping, and data quality metrics. Our work avails the industrial scalability afforded by the Unstructured Information Management Architecture (UIMA) from IBM Watson Research labs, the same framework which underpins the Watson Jeopardy demonstration. This descriptive paper outlines our present work and achievements, and presages our trajectory for the remainder of the funding period. The project is one of the four Strategic Health IT Advanced Research Projects (SHARP) projects funded by the Office of the National Coordinator in 2010.

Session Title: S75 - Papers: A Common Language: Interoperability of Health Data
Session Type: Paper
Location: Gunston (Washington Hilton)
AMIA-0116-A2011. ADEpedia: A Scalable and Standardized Knowledge Base of Adverse Drug Events Using Semantic Web Technology

G. Jiang[1]; H. Solbrig[2]; C. Chute[3]
1. Mayo Clinic College of Medicine, Rochester, MN, United States.
2. Mayo Clinic, , United States.
3. Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, United States.

A source of semantically coded Adverse Drug Event (ADE) data can be useful for identifying common phenotypes related to ADEs. We proposed a comprehensive framework for building a standardized ADE knowledge base (called ADEpedia) through combining the ontology-based approach with the semantic web technology. The framework comprises four primary modules: 1) an XML2RDF transformation module; 2) a data normalization module based on NCBO Open Biomedical Annotator; 3) a RDF store based persistence module; and 4) a front-end module based on a Semantic Wiki for the review and curation. A prototype is successfully implemented to demonstrate the capability of the system to integrate multiple drug data and ontology resources and open web services for the ADE data standardization. A preliminary evaluation is performed to demonstrate the usefulness of the system, including the performance of the NCBO annotator. In conclusion, the semantic web technology provides a highly scalable framework for ADE data source integration and standard query service.

5:15 PM

Session Title: Poster Session 2
Session Type: Poster
Location: Columbia Hall

AMIA-0118-A2011. A Formal Representation of Phenotyping Algorithm Elements

G. Jiang1; J. Pathak2; C. Tao3; H. Solbrig2; C. Chute4
1. Mayo Clinic College of Medicine, Rochester, MN, United States.
2. Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, United States.
3. Health Sciences Research, Mayo Clinic, Rochester, MN, United States.
4. Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, United States.

The objective of the study is to achieve a formal representation model (called PhenoTYPE model) of phenotyping algorithm elements targeting on the collaborative phenotype authoring. We adopted a bottom-up approach by analyzing the structure, elements and authoring workflow process of the existing eMERGE phenotyping algorithms, combining with a top-down approach. We also made a preliminary alignment with the ISO11179 meta-data model, the Clinical Element Models (CEM) and the CDISC Protocol Representation Model (PRM).

AMIA-0274-A2011. The Linked Clinical Data Project: Applying Semantic Web Technologies for Phenomics using Electronic Medical Records

J. Pathak[1]; R. Kiefer[1]; C. Chute[2]
1. Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, United States.
2. Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, United States.

Systematic study of clinical phenotypes is important to better understand the genetic basis of human diseases and more effective gene-based disease management. The Linked Clinical Data (LCD) project at Mayo Clinic aims to develop a semantics-driven framework for high-throughput phenotype extraction, representation, integration, and querying from electronic medical records using emerging Semantic Web technologies, such as Linked Open Data. This poster abstract provides a brief background and overview of the recently initiated LCD project.

Wed, Oct 26 -

8:30 AM

Session Title: S82 - Panel: SNOMED CT Quality Assurance: What's Critical for Users?
Session Type: Panel
Location: Jefferson West (Washington Hilton)

AMIA-0103-A2011. SNOMED CT Quality Assurance: What's Critical for Users?

A. Rector[1]; M. Musen[2]; S. Huff[3]; O. Bodenreider[4]; J. Geller[5]
1. School of Computer Science, University of Manchester, Manchester, United Kingdom.
2. Center for Biomedical Informatics Research, Stanford University, Stanford, CA, United States.
3. Informatics, Intermountain Healthcare, Murray, UT, United States.
4. National Institutes of Health, U.S. National Library of Medicine, Bethesda, MD, United States.
5. NJIT, , United States.

SNOMED CT is now mandated for use in the US and in several other countries. However, SNOMED CT is known to contain errors of various sorts, particularly in the published hierarchies, and to lack any standard binding to Electronic Health Record data. Potentially such errors and lack of standardization make retrieval, post-coordination, and formulation of generic rules for patient care unreliable. How serious are these issues in practice? Now? In the future? What should be done about them? Do recent advances description logic technologies and other techniques help? What are the priorities? Might the UMLS Core Problem Subset provide a focus? Other subsets or modules? What is already under way in the SNOMED organisation, the IHTSDO? This panel will explore these issues in relation to practical applications and standards development. The goal is to stimulate a wider discussion of the priorities for quality assurance of SNOMED CT to achieve meaningful use.

Session Title: S87 - Papers: Medication Reconciliation
Session Type: Paper
Location: Lincoln West (Washington Hilton)

AMIA-0155-A2011. Application of a Temporal Reasoning Framework Tool in Analysis of Medical Device Adverse Events

K. Clark[1]; C. Tao[2]; D. Sharma[2]; C. Chute[3]
1. Department of Biomedical Engineering, University of Minnesota, , MN, United States.
2. Health Sciences Research, Mayo Clinic, Rochester, MN, United States.
3. Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, United States.

The Clinical Narrative Temporal Relation Ontology (CNTRO) project offers a semantic-web based reasoning framework, which represents temporal events and relationships within clinical narrative texts, and infer new knowledge over them. In this paper, the CNTRO reasoning framework is applied to temporal analysis of medical device adverse event files. One specific adverse event was used as a test case: late stent thrombosis. Adverse event narratives were obtained from the Food and Drug Administration's (FDA) Manufacturing and User Facility Device Experience (MAUDE) database. 15 adverse event files in which late stent thrombosis was confirmed were randomly selected across multiple drug eluting stent devices. From these files, 81 events and 72 temporal relations were annotated. 73 temporal questions were generated, of which 65 were correctly answered by the CNTRO system. This results in an overall accuracy of 89%. This system should be pursued further to continue assessing its potential benefits in temporal analysis of medical device adverse events.