Information Based Medicine

The Mayo Clinic/IBM Collaboration

Jeff Tenner tenner@us.ibm.com
Rick Stevens rstevens@us.ibm.com
STSM, IBM Healthcare and Life Science Development
Trademarks and Disclaimers

- © Copyright International Business Machines Corporation 2005
- References in this document to IBM products or services do not imply that IBM intends to make them available in every country.
- The following terms are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both:
- DB2, Websphere,

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

- Other company, product and service names may be trademarks or service marks of others.
- Information is provided "AS IS" without warranty of any kind.
- All customer examples described are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics may vary by customer.
- Information in this presentation concerning non-IBM products was obtained from a supplier of these products, published announcement material, or other publicly available sources and does not constitute an endorsement of such products by IBM. Sources for non-IBM list prices and performance numbers are taken from publicly available information, including vendor announcements and vendor worldwide homepages. IBM has not tested these products and cannot confirm the accuracy of performance, capability, or any other claims related to non-IBM products. Questions on the capability of non-IBM products should be addressed to the supplier of those products.
- All statements regarding IBM future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only. Contact your local IBM office or IBM authorized reseller for the full text of the specific Statement of Direction.
- Some information in this presentation addresses anticipated future capabilities. Such information is not intended as a definitive statement of a commitment to specific levels of performance, function or delivery schedules with respect to any future products. Such commitments are only made in IBM product announcements. The information is presented here to communicate IBM's current investment and development activities as a good faith effort to help with our customers' future planning.
Agenda

- Objectives of collaboration
- Solution Technology and Architecture
  - Data Discovery and Query Builder
  - Text Analysis
  - Microarray Data
  - Future
Enable non-IT proficient researchers and physicians to build complex queries without dealing with the complexities of the source data, database schema, SQL, or relational concepts.

- Ability to define co-existent data elements
- Durable
  - Yesterday’s queries work with the rapidly evolving data model of today and tomorrow.
- Security and auditability
  - By role authorization
  - By data content
- Data from multiple sources – internal and external
  - Must support warehouses and federation
- Based on standards
- Handles unstructured text
- Integrates with partner tools and applications
- Thin client web-browser based
An example query:

- **Find**
  - Patients diagnosed with acute myeloid leukemia (AML) as compared to infectious mononucleosis
  - Male
  - Living
  - Diagnosed while less than 60 years of age
  - Diagnosed while living in Illinois
  - Normal white blood cell count
  - Mention of prednisone treatment in clinical notes
  - For which good quality microarray expression data exists
Mayo Clinic Clinical Genomics Architecture

Gene Expression
Clinical Notes
Registration, Diagnoses Codes and others
Laboratory Results and Reports

Web services
Genes@Work
Affymetrix
BioConductor
Spotfire
PubMed
R
SAS
Mayo Clinic

DB2

Websphere

XML transform
ETL
GoldenGate Replication

Document Driver

APIs

Annotators
CAS Extraction
UIMA

DB2 Warehouse

Researchers

Search, Analyze, Export Data and Results
Data Discovery and Query Builder

1. Retrieve Fields
2. Create Query
3a. Logical-to-Physical Mapping
3. Generate Concrete Query
4. Execute Query
5. Return Query Results
6. Analysis

Generated UI

Query Abstraction Model

Data Abstraction Model

Query Translation and Execution

Physical Data Repository

Studio
Data Discovery and Query Builder
Integration Points

- **Web Services**
  - Remote query execution

- **Validator plug-in**
  - User authentication

- **Authorization plug-in**
  - Control of DAM view based on user authority

- **Skin plug-in**
  - User interface appearance customization

- **Auditor plug-in**
  - Tracking execution of queries

- **Field Features**
  - Term expansion, text search, etc.

- **API Programmatic access to DDQB**

- **Output Formatter plug-in**
  - Editing and selection of query results

- **Data Abstraction Model**
  - Define logical to physical mapping and drive UI

- **Analysis plug-in**
  - Launch analysis routine and pass query results

- **IBM Life Sciences**

---

8
Text Analysis

IBM has worked with the Mayo Clinic to secure data for this demonstration that is either wholly fictional or has been modified by law to ensure anonymity.
Clinical Note (Document ID is 72432230)

Reason for Visit

outpatient note. The patient returns to the Hospital. He is post consolidation chemotherapy with high-dose ara-C for his AML-M4.

History of Present Illness

Please see detailed note from October. The patient continues to do well as an outpatient. He denies any mouth pain—no nausea or vomiting. He has been eating and drinking without any difficulty. He denies any chest pain or shortness of breath. The patient did complete his prednisone eye drops yesterday, and he denies any eye discomfort.
Clinical Note (Document ID is 72432230)

Reason for Visit

outpatient note. The patient returns to the Hospital. He is post consolidation chemotherapy with high-dose ara-C for his AML-

History of Present Illness

Please see detailed note from October. The patient continues to do well as an outpatient. He denies any mouth pain--no nausea or vomiting. He has been eating and drinking without any difficulty. He denies any chest pain or shortness of breath. The patient did complete his prednisone eye drops yesterday, and he denies any eye discomfort.
Clinical Note (Document ID is 72432230)

Reason for Visit

outpatient note. The patient returns to the Hospital. He is post consolidation chemotherapy with high-dose ara-C for his AML-M4.

History of Present Illness

Please see detailed note from October. The patient continues to do well as an outpatient. He denies any mouth pain—no nausea or vomiting. He has been eating and drinking without any difficulty. He denies any chest pain or shortness of breath. The patient did complete his prednisone eye drops yesterday, and he denies any eye discomfort.
Clinical Note (Document ID is 72432230)

Reason for Visit

outpatient note. The patient returns to the Hospital. He is post consolidation chemotherapy with high-dose ara-C for his AML-M4.

History of Present Illness

Please see detailed note from October. The patient continues to do well as an outpatient. He denies any mouth pain--no nausea or vomiting. He has been eating and drinking without any difficulty. He denies any chest pain or shortness of breath. The patient did complete his prednisone eye drops yesterday, and he denies any eye discomfort.
Clinical Note (Document ID is 72432230)

Reason for Visit

outpatient note. The patient returns to the Hospital. He is post consolidation chemotherapy with high-dose ara-C for his AML-M4.

History of Present Illness

Please see detailed note from October. The patient continues to do well as an outpatient. He denies any mouth pain—no nausea or vomiting. He has been eating and drinking without any difficulty. He denies any chest pain or shortness of breath. The patient did complete his prednisone eye drops yesterday, and he denies any eye discomfort.
Clinical Note (Document ID is 72432230)

Reason for Visit

outpatient note. The patient returns to the Hospital. He is post consolidation chemotherapy with high-dose ara-C for his AML-M4.

History of Present Illness

Please see detailed note from October. The patient continues to do well as an outpatient. He denies any mouth pain—no nausea or vomiting. He has been eating and drinking without any difficulty. He denies any chest pain or shortness of breath. The patient did complete his prednisone eye drops yesterday, and he denies any eye discomfort.
Clinical Note (Document ID is 72432230)

Reason for Visit

outpatient note. The patient returns to the Hospital. He is post consolidation chemotherapy with hogh-dose ara-C for his AML-M4.

History of Present Illness

Please see detailed note from October. The patient continues to do well as an outpatient. He denies any mouth pain--no nausea or vomiting. He has been eating and drinking without any difficulty. He denies any chest pain or shortness of breath. The patient did complete his prednisone eye drops yesterday, and he denies any eye discomfort.
Example Text Analysis
Annotator Topology

Annotators:
- Initialize/Detagging
- Context Free Tokenizer
- Sentence Detector
- Context Sensitive Tokenizer
- Abbreviation Disambiguation
- Context Sensitive Spell Correction
- Lexical Normalization
- Part-of-Speech Tagger
- Shallow Parser
- Drug Annotator (Dictionary)
- Disorders Annotator (Dictionary)
- Findings Annotator (Dictionary)
- SAIL Disorder
- SAIL Signs/Symptoms
- SAIL Drug
- Negation

Resources:
- Shallow Parse Rules
- SAIL Models
- Part-of-Speech Model
- Spelling Lexicon
- Abbreviation Lexicon
- Negation Rule Base
- Spell Correction Model
- Sentence Detection Model
- Abbreviation Model
- Stopword Lexicon
- Part-of-Speech Lexicon
- Norm Lexicon
- Findings Dictionary
- Diagnosis Dictionary
Microarray Data Flow

- **Web Interface**
- **Microarray Labs**
  - **Values**
  - **Normalized Values**
    - FTP
  - **Raw Image**
    - FTP
  - **Chip Definition**
    - FTP

- **Java App**
  - **MAGEstk**
    - Temporary data file
    - Java API from IBM

- **Microarray Data Warehouse**
  - MAGE tables
  - External files (locations encoded in MAGE-ML file)
    - Java API from IBM

- **Web Interface**
  - Analysis routines
  - Normalized Expression values

- **MAGE-ML**
An example query:

- **Find**
  - Patients diagnosed with acute myeloid leukemia (AML) as compared to infectious mononucleosis
  - Male
  - Living
  - Diagnosed while less than 60 years of age
  - Diagnosed while living in Illinois
  - Normal white blood cell count
  - Mention of prednisone treatment in clinical notes
  - For which good quality microarray expression data exists
Current Conceptual Architecture

EMR
Demographics
Coded Diagnostics
Clinical Notes
Gene Expression

Medical Information Repository
Future Problem: Separate EMRs at Each Site

Solution: Use Standards-Based EMR Model
Future Problem: Separate EMRs at Each Site
Solution: Use Standards-Based Data Model

- EMR
- EMR
- EMR
- Demographics
- Coded Diagnostics
- Clinical Notes
- Gene Expression

Medical Information Repository
Move from this...

IBM Life Sciences

Gene Expression
Clinical Notes
Registration, Diagnoses Codes and others
Laboratory Results and Reports

Gene Expression
Clinical Notes
Registration, Diagnoses Codes and others
Laboratory Results and Reports

Genes@Work
Affymetrix
BioConductor
Spotfire
PubGene
R
SAS
Mayo Clinic

DB2

Websphere

XML transform
ETL
GoldenGate Replication

CDA / XML
MAGE-ML

Document Driver

APIs

Annotators
CAS Extraction

Web services

Researchers
Search, Analyze, Export Data and Results

DB2 Warehouse
IBM ^ pSeries
New Bioinformatic Tools and Techniques:

- Architecture built on industry standards
- Innovative data abstraction layer
- Incorporates advanced text analysis
- Designed for resilience and customization
- Integrated with partner applications for a complete solution
The End