Content-based Image Tagging and Online Image Database System

Student: Huanmei Wu (NU)
Advisors: Prof. David Kaeli (NU)
Prof. Betty Salzberg (NU)
Contributors: Becky Norum (NU)
Pat Muraca (Clinomics)

This work was supported in part by CenSSIS, the Center for Subsurface Sensing and Imaging Systems, under the Engineering Research Centers Program of the National Science Foundation (Award Number EEV-9986821).
Goals

- Set up an online image database for Clinomics that enables fast searching and querying images based upon image contents

**Desired Characteristics:**

- Content-based image tagging using XML adopting MPEG-7 standard
- Relational database enables complex online queries
- Enable web database security and layered access rights of users
- Efficient search and query of image content and metadata
- Easy upload and registration of user images and metadata
- Indexing and partitioning of the database for better performance
Significance

* Great importance for clinicians and researchers

* Aid the clinicians in diagnostic assessment, therapeutic approach and prognostic evaluation

* Provide researchers with enhanced means of gathering, assessing and collating relevant clinical data
Related Work

• Information retrieval databases coupled to database search system exist (such as NCBI, PubMed, MMDB, etc), but do not provide tissue standards, patient information or relate an image with its associated patient history.

• Some tissue banks provide both tissue samples and computer accessible information about the tissues they bank. However, they do not provide a convenient way to access both tissue information and molecular data relating to a particular tissue source in one location. Also, existing data banks do not provide relational searching capabilities and feature tagging strategies.

• Databases that catalog medical findings into an online format provide an interface displaying links that identify particular cases assessed by pathology departments, but do not provide a way to correlate new data with the existing data, or to identify relationships between biological characteristics of the tissue samples and multiple patients.
Technical Approach I – Image Tagging

Requirements:
• Perform content query within images
• Perform relational operations upon features present in a single image

Solution:
Use XML to represent structure and hierarchy in a single data file including image content, patient history and other metadata while adopting some MPEG-7 standards

```xml
<Image>
  <!-- General Cell Information -->
  <CellInformation>
    <ID> 9 </ID>
    <ClinomicsID> 931175495 </ClinomicsID>
    <DOB> 2/7/30 </DOB>
    <SEX> F </SEX>
    <COLL_DATE> 11/2/1993 </COLL_DATE>
    <Primary_site> Breast </Primary_site>
    <INITIAL> II </INITIAL>
    <GRADE> POORLY DIFFERENTIATED </GRADE>
    <HISTOLOGY> UNKNOWN </HISTOLOGY>
    <PRIM_SITE2> NONE </PRIM_SITE2>
    <PRIM_DATE> 4/1/1992 </PRIM_DATE>
    <MET1_SITE> NONE </MET1_SITE>
    <MET1_DATE> NONE </MET1_DATE>
    <TUBE_TYPE> p </TUBE_TYPE>
  </CellInformation>
  <StillRegion id="IMG0001">
    <MediaProfile>
      <MediaFormat>
        <FileFormat>jpeg</FileFormat>
        <System>PAL</System>
        <Medium>CD</Medium>
        <Color>color</Color>
        <FileSize>332.228</FileSize>
      </MediaFormat>
      <MediaCoding>
        <Resolution>200</Resolution>
        <FrameWidth>1165</FrameWidth>
        <FrameHeight>783</FrameHeight>
        <CompressionFormat>JPEG</CompressionFormat>
      </MediaCoding>
      <Identifier IdOrganization='Clinomics' IdName='BreastCancerCell'>BreastCancerCell//image0001</Identifier>
      <Locator>
        <MediaURL>file://D:/Breast/cells/imag0001.jpg</MediaURL>
      </Locator>
    </MediaProfile>
    <StructuredAnnotation>
      <Who>Jeanne Becker</Who>
      <whatObject>Human primary breast tumor cells</whatObject>
      <WhatAction> growing in a NASA Bioreactor </WhatAction>
      <where> St. mary Hospital </where>
      <When> 09/25/2000 </When>
      <why> Investigate tumor cells behaviour on microcarrier beads </why>
      <TextAnnotation xml:lang='en-us'>Higher magnification of view illustrating breast cancer cells with intercellular boundaries on bead surface</TextAnnotation>
    </StructuredAnnotation>
  </StillRegion>
</Image>
```
Technical Approach II:
Web-based Database Application

How does this work?
1) Client sends a request to the web server (e.g., a database search request).
2) Web server directs request to the appropriate servlet.
   (A *servlet* is a piece of Java code that executes within a web server.)
3) The servlet processes the request, generates the appropriate query and queries
   the Oracle database using *JDBC* (the Java interface to database systems).
4) The database server generates a set of results and returns it to the servlet.
5) The servlet then processes the response for display in a browser window.
6) The web server sends the response in HTML form to the client browser.

**Requirements:**
- Application can be accessed from a variety of platforms and browsers
- Application can deliver dynamic web pages generated through efficient database queries

**Solution:**
- Use 3-Tier Client/Server Model
Current Status

- Application prototype in development phase
- Strategy to optimize database performance under consideration

Industrial Linkages

Clinomics BioSciences, an experienced and acknowledged leader in tissue products and services, has offered clinical images and associated metadata to develop a prototype XML-enabled database system.
Future Plans

- Develop a web-based application that can initially:
  1) Upload and download images and their associated metadata.
  2) Provide comprehensive searching capabilities.
- Refine system to allow:
  1) Content-based indexing and retrieval using XML tags.
  2) A graphical interface to allow researchers to query parts of any image.

References:


Contact Info:

Name: Huanmei Wu
Institution: Northeastern University
Email: hwu@ece.neu.edu