Announcements

- VisualStudio Express
  - Free, Hobbyist version of VS 2005
- Presentations, Tuesday Jan 23
  - 15 minute presentation + 3 minutes discussion
  - PowerPoint slides
  - Group order: A, B, C, D

Today’s lecture

- Handwriting Recognition
- Structure Recognition
- Classification
- Annotation
- JNT Note Format

Ink Analysis for Search

- Output
  - Mapping of search results to source
  - Reflect underlying structure
  - Handle different types of search queries
    - Raw text
    - Boolean
    - Typed queries (“481 as a course number”)
    - Object queries (Course numbers)
    - Environment (List, Prose, Mathematics, . . . )

Ink Analysis Pipeline

Filter
Structure
Recognize
Classify
Annotate

Handwriting Recognition:
Identify the following words

- programmer
- optimistic
- data
- survey
- structure
Recognition results

Recognizer Architecture

Recognizer Training
- Collect large set of training data
- Samples of known inputs that can be used to set “weights” in reco engine
- Needed to build a recognizer
  - Dictionary
  - Language samples
- Commercial recognizers based on massive data sets

Tablet PC Recognition API
- Basic idea:
  - Ink In, Text Out

Recognition Code I
```csharp
private Recognizers recognizers;
private Recognizer recognizer;

public Form1() {
    InitializeComponent();
    this.inkCollector = new InkCollector(this.inkPanel.Handle);
    this.inkCollector.Enabled = true;
    this.recognizers = new Recognizers();
    this.recognizer = recognizers.GetDefaultRecognizer();
}
```

Recognition Code II
```csharp
private void OnRecoClick(object sender, EventArgs e) {
    RecognizerContext recoContext = this.recognizer.CreateRecognizerContext();
    recoContext.Factoid = GetFactoid();
    recoContext.Strokes = this.inkCollector.Ink.Strokes;
    recoContext.EndInkInput();
    RecognitionStatus recoStatus;
    RecognitionResult recoResult = recoContext.Recognize(out recoStatus);
    if (recoStatus != RecognitionStatus.NoError)
        return;
    string result = recoResult.TopString;
    RecognitionAlternate topAlt = recoResult.TopAlternate;
```
Factoids
- Bias the recognizer towards certain types of content
  - DEFAULT
  - CURRENCY
  - NUMBER
  - TELEPHONE
  - EMAIL
  - UPPERCCHAR

Reading Journal Notes
- Journal Reader to import .JNT
- .JNT -> XML -> Custom Format
- Journal format gives an initial parsing
  - You may want to undo this parsing and work with ink at the page level

JNT Format
- Journal Document
  - List of Journal Pages
- Journal Page
  - List of Content
- Content
  - Journal Drawing, Journal Paragraph, other stuff

Journal Drawing
- Uninterpreted Ink
  - Base64String

if (childNode.Name.ToLower().Equals("inkobject")) {
    string base64Ink = childNode.InnerText;
    ink = new Ink();
    ink.Load(Convert.FromBase64String(base64Ink));
}

Text Structure
- JournalParagraph
  - List of JournalLines
- JournalLine
  - List of JournalInkWords
- JournalInkWord
  - Alternate List
  - Uninterpreted Ink

Shape recognition
- Surprisingly challenging because of drawing artifacts
  - Open figures
  - Multiple strokes
  - Imprecise corners
  - Arrows
Structure recognition

Basic approach for structure recognition
- Grouping by rectangular region
- Heuristics for separating regions
  - White space
  - Separating lines

General approach to recognition/classification
- Extraction of features
- Objects become points in high dimensional space
- Construct mapping from features to classes

Clustering

Learning

Heuristic
- Programmatically determine classification based on features
Classification

- Identify different types of text
  - Mathematics
  - Prose
  - Lists
  - Brainstorming
  - Code
  - Domains
    - Chemistry, Physics, Algorithms,

Annotation

- Identify annotation marks
  - Highlighted text
  - Circles
  - Check marks
  - Cross out