# Investigating the Characteristics of Mathematical Notations

**Clare So**
clare@scl.csd.uwo.ca

**Stephen Watt**
watt@scl.csd.uwo.ca



ORCCA Joint Lab Meeting. UWO. April 8, 2005.

# Why Studying Math Notations?

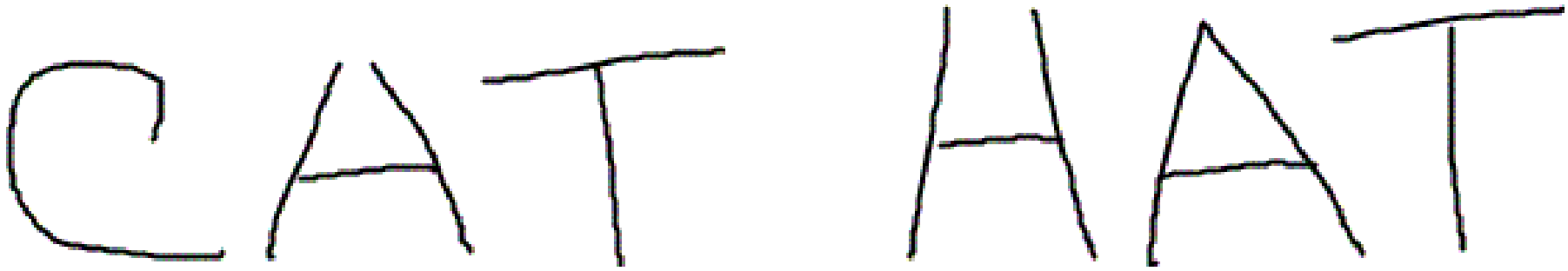
- Guide the mathematical expressions handwriting recognizer to eliminate results of recognition that does not make sense
- Preserve semantical information of expressions when translating between TEX and MathML

# Math Notation is a Language

- A natural language have certain *patterns* of usages
  - For example, “x” and “z” are not letters commonly used in English words
- Usages of math notation mainly based on conventions
  - $f$ and $g$ for functions
  - $i$ and $j$ for integers
  - $A$ and $B$ for sets
  - $i$ for $\sqrt{-1}$
  - ...

# Previous Knowledge of Language Important

- Why we recognize the words “cat” and “hat” although “H” and “A” are writen similarly?
- In natural language handwriting recogition, a built-in dictionary is used to eliminate results that does not make sense



# How About Math?

- We need a dictionary for math in a handwriting recognizer
- Not many studies related to math “linguistics”
- This project provides a foundation of building a math dictionary

# Data Collection

- All mathematical articles in ArXiv database from 2000 to 2005 are collected
- Articles are put into their repective top-level *Mathematical Subject Classification* (2000). There are 63 classification in total representing a field in advanced mathematics.
- Math expressions are extracted from articles TeX source and translated to MathML

# Analyzing Each Classification

- We keep track of the following information in each classification
  - Common expressions
  - Frequency of single character symbols
    - Frequency of identifiers
    - Frequency of different operators
  - Patterns of expressions
- Histrograms are to be built for keeping track of symbols' frequencies

# Generating Patterns by Antiunification

- We want to recognize commons parts in expressions
  - Let's say we have expressions $F(x, G(x, y), z)$ and $F(a, G(a, b), z)$. It would be useful to obtain the pattern $F(\alpha_1, G(\alpha_1, \alpha_2), z)$
- Concept of antiunification was first discussed by Robinson in 1965
- Algorithm for generating such patterns (aka antiunifiers) was discovered by Plotkin in 1970
- We are performing antiunification to MathML expression trees
  - For example, $x^1$, $x^2$ and $x^3$ all have $x^\alpha$ as the antiunifier

# Results

- The most common identifiers in classifications:
  - 08 (General algebraic systems):
    $n \quad A \quad , \quad a \quad i$
  - 12 (Field theory and polynomials):
    $, \quad n \quad x \quad i \quad K$
  - 31 (Potential theory):
    $x \quad , \quad n \quad z \quad d$
  - 42 (Fourier analysis):
    $, \quad n \quad x \quad k \quad j$
  - 83 (Relativity and gravitational theory):
    $, \quad i \quad x \quad M \quad t$

# Results

- The most popular expressions in classifications:
  - 08 (General algebraic systems):

```
<mrow><mi>A</mi></mrow>
<mrow><mo>-</mo><mn>1</mn></mrow>
<mrow><mi>G</mi></mrow>
<mtd/>
<mrow><mo>(</mo><mi>X</mi><mo>)</mo></mrow>
```

  - 12 (Field theory and polynomials):

```
<mrow><mo>(</mo><mi>x</mi><mo>)</mo></mrow>
<mrow><mo>-</mo><mn>1</mn></mrow>
<mrow><mi>K</mi></mrow>
<mrow><mi>p</mi></mrow>
<mrow><mi>n</mi><mo>-</mo><mn>1</mn></mrow>
```

# Results

- The most popular expressions in classifications:
  - 31 (Potential theory):

```
<mrow><mo>(</mo><mi>x</mi><mo>)</mo></mrow>
<mrow><mo>(</mo><mi>z</mi><mo>)</mo></mrow>
<mtd/>
<mrow><mo>-</mo><mn>1</mn></mrow>
<mrow><mo>(</mo><mi>r</mi><mo>)</mo></mrow>
```

  - 42 (Fourier analysis):

```
<mrow><mo>(</mo><mi>x</mi><mo>)</mo></mrow>
<mrow><mo>-</mo><mn>1</mn></mrow>
<mtd/>
<mrow><mo>(</mo><mi>z</mi><mo>)</mo></mrow>
<msup><mi>L</mi><mn>2</mn></msup>
```

# Results

- The most popular expressions in classifications:
  - 83 (Relativity and gravitational theory):

```
<mtd/>
<mrow><mo>-</mo><mn>1</mn></mrow>
<mrow><mo>(</mo><mi>t</mi><mo>)</mo></mrow>
<mrow><mo>(</mo><mi>x</mi><mo>)</mo></mrow>
<mrow><mo>(</mo><mn>1</mn><mo>)</mo></mrow>
```

# Results

- Some antiunifiers:
  - 00 (General):

```
<mfrac><mi></mi><orccaPattern index="0"/></mfrac>
<mfrac><mi>k</mi><orccaPattern index="0"/></mfrac>
<mfrac><mn>1</mn><mrow><mn>1</mn><mo>-</mo><orccaPattern index="0"/>
     </mrow></mfrac>
<mfrac><mn>1</mn><mrow><mn>5</mn><mo></mo><mn>1</mn><msup><mn>0</mn>
     <orccaPattern index="0"/></msup></mrow></mfrac>
<mfrac><mrow><mi>d</mi><orccaPattern index="0"/></mrow><mrow><mi>d</mi>
     <orccaPat tern index="1"/></mrow></mfrac>
<mfrac><mn>25</mn><orccaPattern index="0"/></mfrac>
```