

Search and Inference with Diagrams

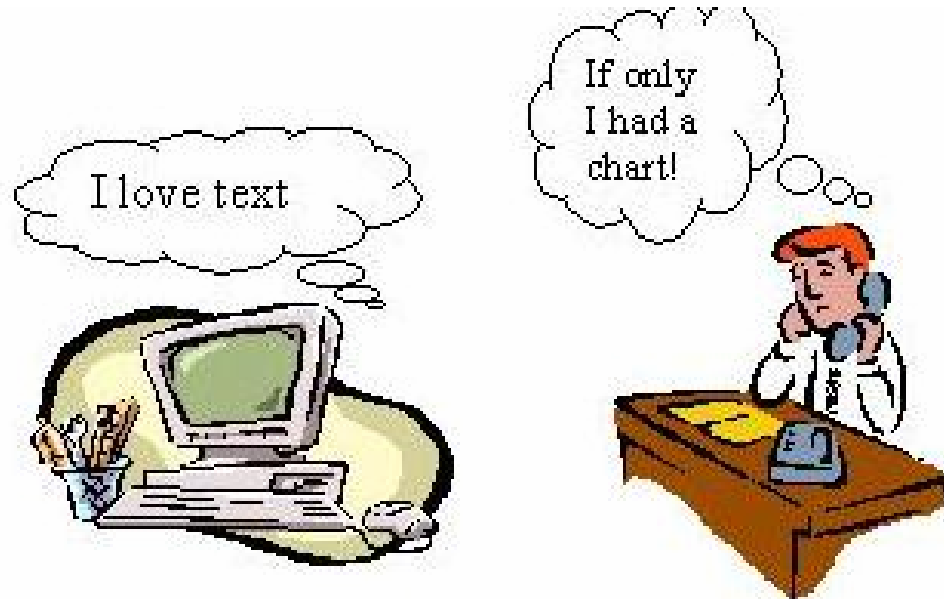
Michael Wollowski, Ph.D.

Associate Professor of Computer Science and Software
Engineering

Rose-Hulman Institute of Technology

Terre Haute, IN 47803, USA

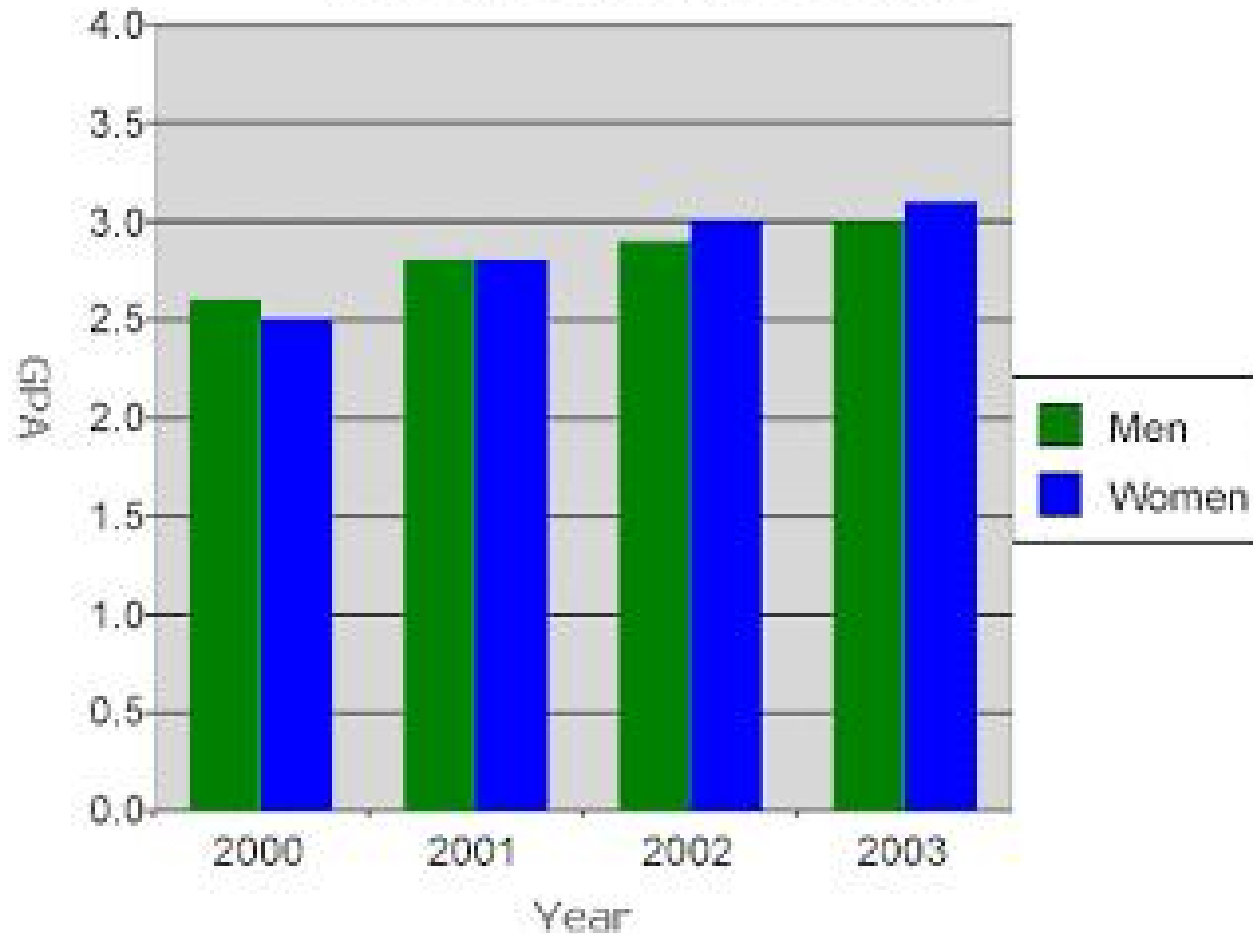
Motivation



- Search engines search text
- People prefer diagrammatic representations
- Diagrams are typically stored as image files
- Information in image files cannot be searched

Sample Diagram

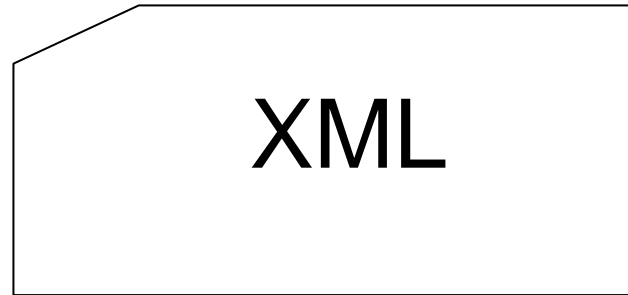
4-Year GPA by Gender



Outline

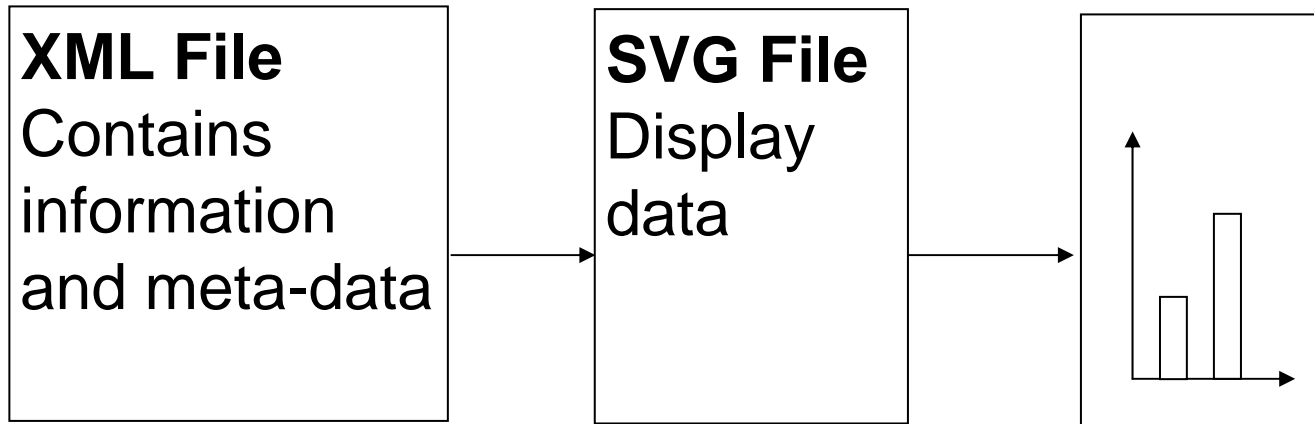
- Approach taken
- Theoretical background
- Sample document
- Search engine
- Inference engine
- Conclusions

Approach



- Represent information in XML format
- Render XML files as charts
- Provide custom-fit search and inference engine

Process



Stored on web server;
accessible to search engine

Server-side processing
(when file is requested)

Client-side display

Benefits of our Approach

- Information is semantically annotated
- Information is accessible to search and inference engines
- Diagrams help people extract information

Theoretical Background

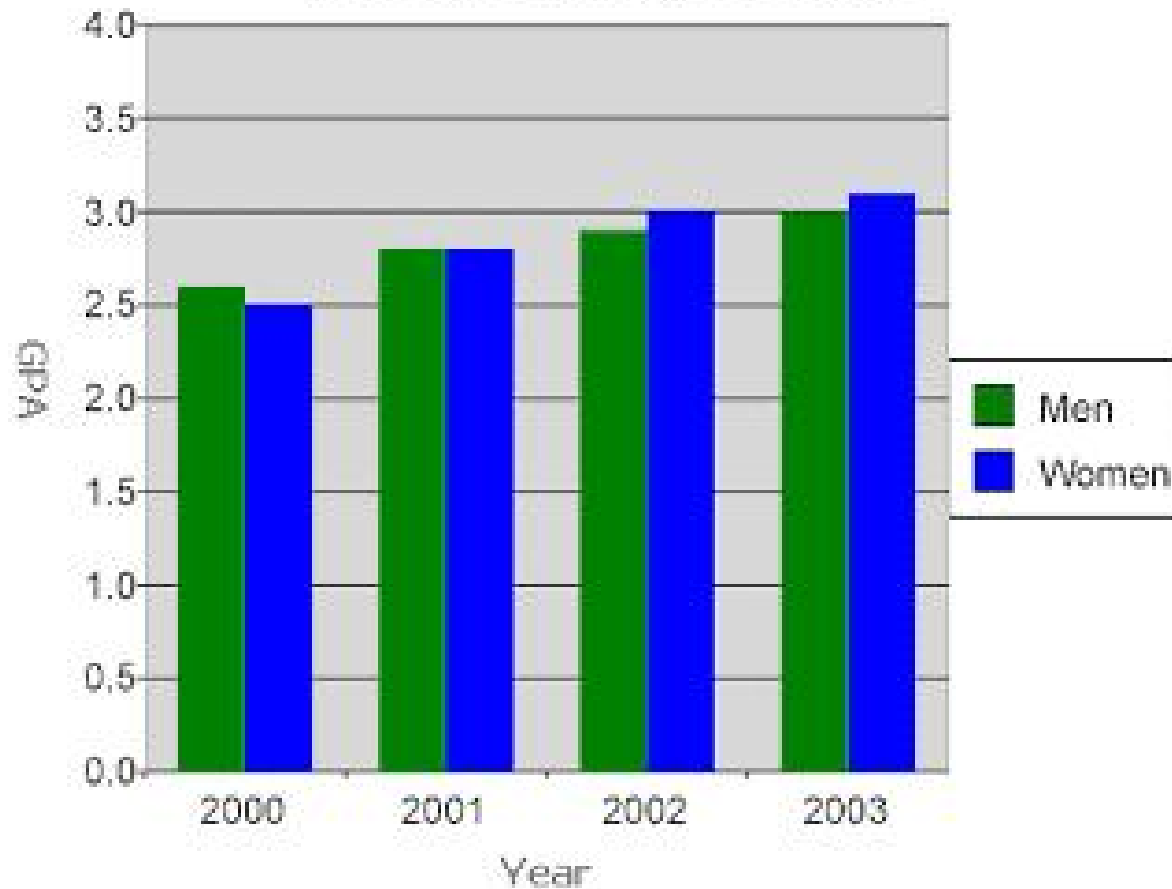
- People successfully use charts to represent and extract information
- Charts are highly structured, as such a syntax can be given for them just like for FOL
- A semantics for charts can be given either directly or by correspondence

Sample XML Document

```
<?xml version="1.0" ?>
<chart type="bar">
  <title>4-Year GPA by Gender</title>
  <legend />
  <plotarea>
    <axis variable="x1">Year</axis>
    <axis variable="x2" min="0" max="4.0"
      step="0.5">GPA</axis>
  </plotarea>
  <dataset label="Men">
    <datapoint x1="2000" x2="2.6" />
    <datapoint x1="2001" x2="2.8" />
    <datapoint x1="2002" x2="2.9" />
    <datapoint x1="2003" x2="3.0" />
  </dataset>
  <dataset label="Women">
    <datapoint x1="2000" x2="2.5" />
    <datapoint x1="2001" x2="2.8" />
    <datapoint x1="2002" x2="3.0" />
    <datapoint x1="2003" x2="3.1" />
  </dataset>
</chart>
```

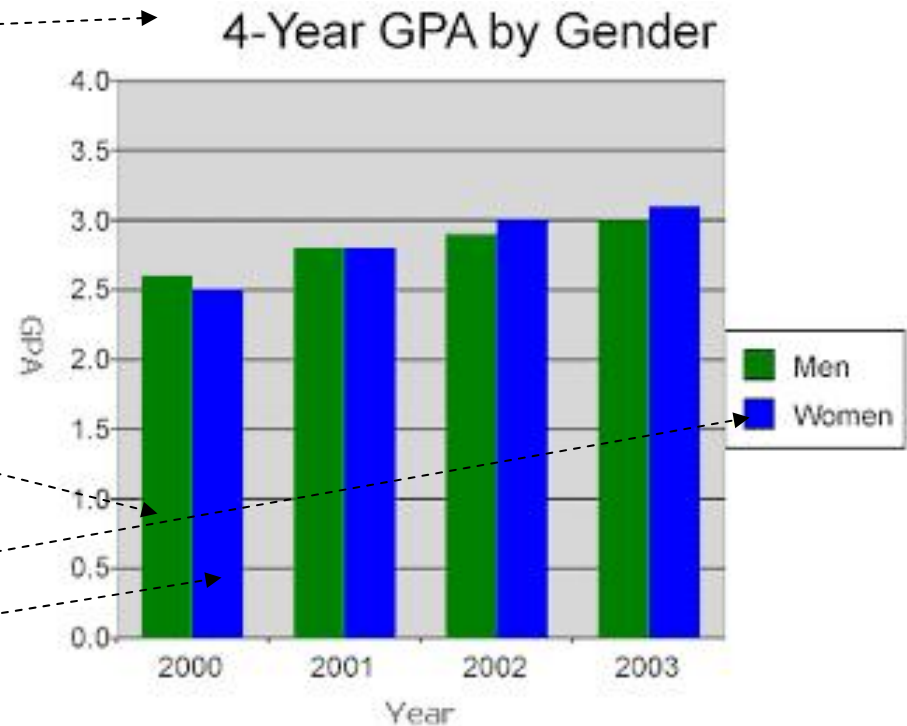
Document Rendered

4-Year GPA by Gender



Correspondence of XML and Diagram

```
<?xml version="1.0" ?>
<chart type="bar">
  <title>4-Year GPA by Gender</title>
  <legend />
  <plotarea>
    <axis variable="x1">Year</axis>
    <axis variable="x2" min="0" max="4.0"
      step="0.5">GPA</axis>
  </plotarea>
  <dataset label="Men">
    <datapoint x1="2000" x2="2.6" />
    <datapoint x1="2001" x2="2.8" />
    <datapoint x1="2002" x2="2.9" />
    <datapoint x1="2003" x2="3.0" />
  </dataset>
  <dataset label="Women">
    <datapoint x1="2000" x2="2.5" />
    <datapoint x1="2001" x2="2.8" />
    <datapoint x1="2002" x2="3.0" />
    <datapoint x1="2003" x2="3.1" />
  </dataset>
</chart>
```



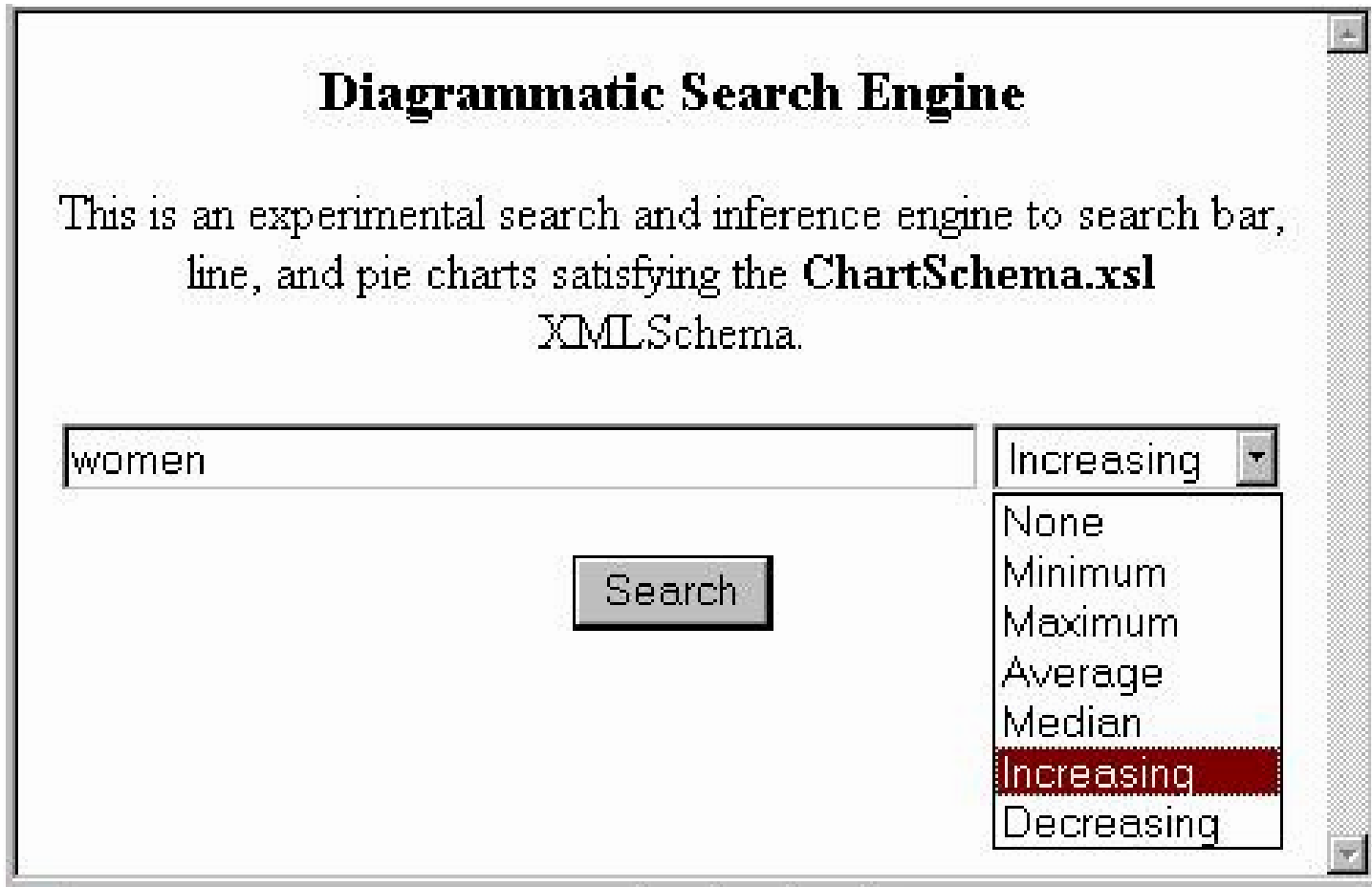
Key XML Components

- `Datapoint` element for each piece of data
- `Dataset` elements enclose & label datapoints
- `Title` and `legend` elements for annotations
- `Plotarea` specifies axis labels and decoration
- `Chart type` element to specify kind of chart (bar, line, or pie chart)

Software Components

- **XML Schema** for 2D bar, line, and pie-charts
- **Servlet** that processes XML files and outputs SVG files
- **XSL stylesheet** for invoking servlet
- *Sample files* containing student information

Interface to Search/Inference Engine



Results Screen

Results for: **women** with inference **Increasing**

1. **4-Year GPA by Gender:** <http://localhost:8080/servlets/xml/barchart.xml>
 - *Women* (Data increases.)
 - Year: 2000 GPA: 2.5
 - Year: 2001 GPA: 2.8
 - Year: 2002 GPA: 3.0
 - Year: 2003 GPA: 3.1
2. **4-Year GPA by Gender:** <http://localhost:8080/servlets/xml/barchart2.xml>
 - *Women* (Data increases.)
 - Year: 2000 GPA: 2.5
 - Year: 2001 GPA: 2.8
 - Year: 2002 GPA: 2.8
 - Year: 2003 GPA: 3.1

Key Ideas

- **Search engine** takes advantage of semantic nature of elements
- It locates explicitly contained information
- **Inference engine** makes available implicitly contained information
- It supports determining the: *minimum*, *maximum*, *median*, and *average* of data that satisfies a query, as well as whether the data *increases* or *decreases*

Software Components

- An XML search and inference engine **servlet**

Conclusions

- We developed a general process of storing diagrammatic information in a sentential format
- Stored information is accessible to inference engines yet rendered as diagram
- Our system is sufficiently general to store information that is typically displayed in bar, line, or pie-charts
- We modified the XML Schema so that data from MS Spreadsheets can be automatically extracted to our XML format

Future Work

- 3D charts
- Determining trends
- Diagrams other than charts