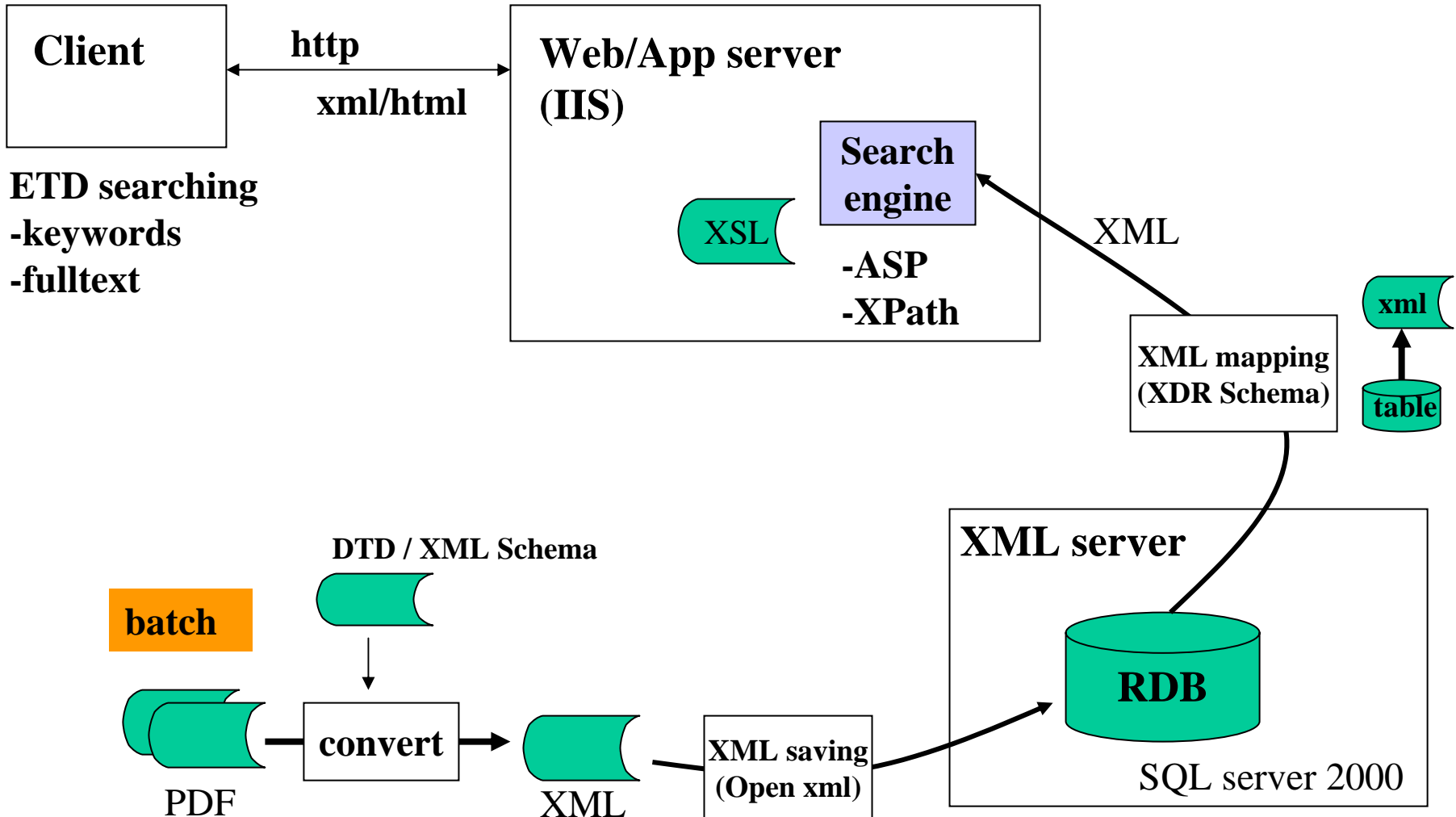


# ETD DTD

- Based on ETD-ML1.1(Virginia Tech)
  - convert SGML DTD to XML DTD
- Added some tags
  - such as TableOfConetents, ListOfPicture,  
ListOfTables
- Added “id” attribute to some tags
  - such as bibliography, vita

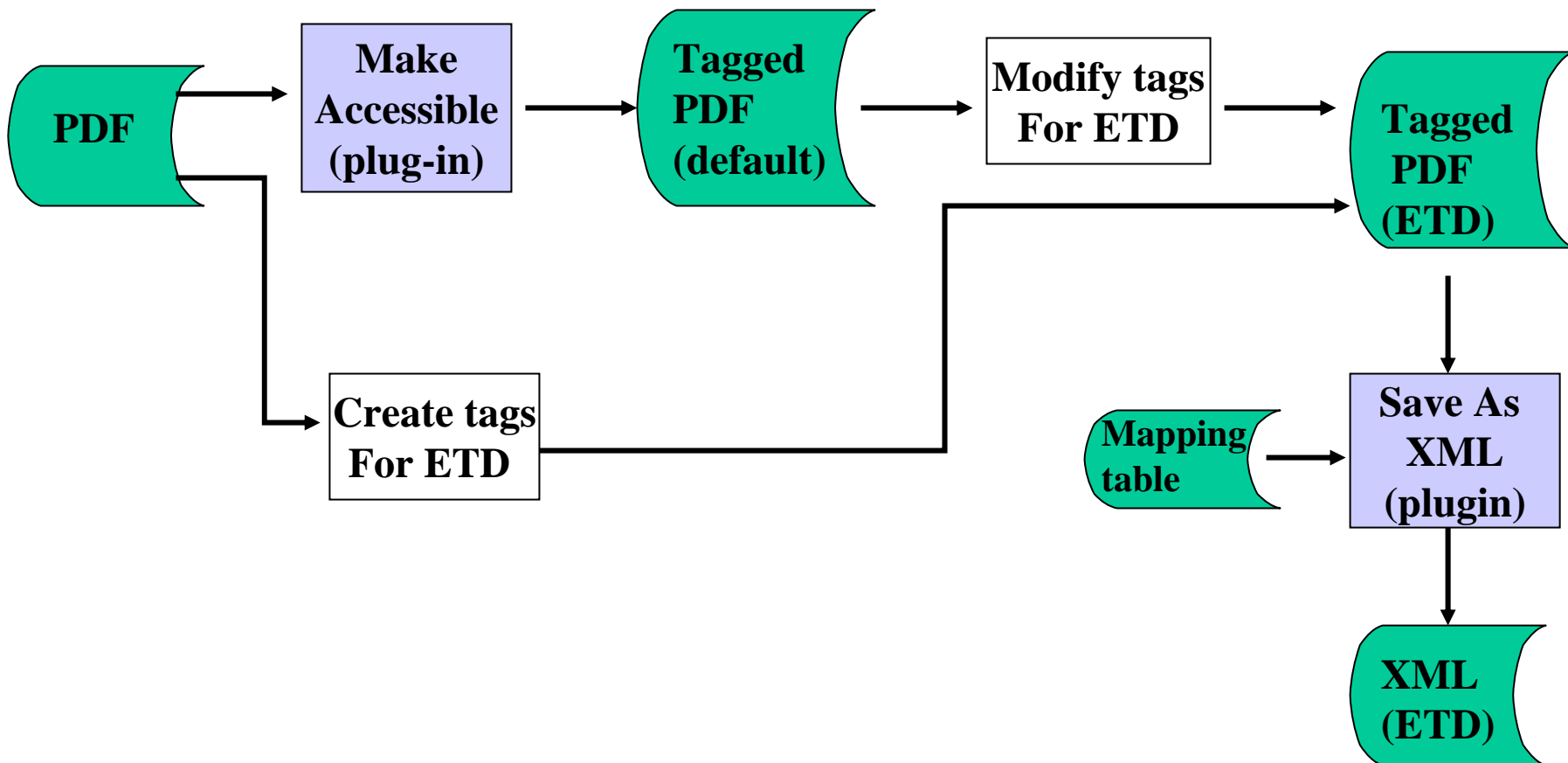
# ETD System Structure

**Online**



# Convert PDF to XML

-- Adobe “save as xml”



# Method1: Create tags for tagged PDF (ETD)

Adobe Acrobat - [TKW01diss3.pdf]

File Edit Document Tools View Window Help

119%

Tags

Tags Root

- <front> front
  - <title> title
    - Strategies for Handli...
  - <author> author
  - <abstract> abstract
    - <p> paragraph
      - Geographic informatio...
- <body> body
  - <chapter> chapter
    - <head> head
      - Chapter 1 Introduction
    - <p> paragraph
      - It is a challenge to ...
    - <p> paragraph
    - <section> section
      - <head> head
        - 1.1 Terminology
      - <p> paragraph
    - <section> section
      - <head> head
        - 1.2 Motivation
      - <p> paragraph
        - This section uses exa...
    - <chapter> chapter
      - <head> head
        - Chapter 2 Data Quali...
  - <back> back
    - <bibliography> bibliography
    - <appendix> appendix

Strategies for Handling Spatial Uncertainty due to  
Discretization

By Thomas K. Windholz

Thesis Advisor: Dr. M. Kate Beard-Tisdale

An Abstract of the Thesis Presented  
in Partial Fulfillment of the Requirements for the  
Degree of Doctor of Philosophy  
(in Spatial Information Science and Engineering)  
May, 2001

Geographic information systems (GISs) allow users to analyze geographic phenomena within areas of interest that lead to an understanding of their relationships and thus provide a helpful tool in decision-making. Neglecting the inherent uncertainties in spatial representations may result in undesired misinterpretations. There are several sources of uncertainty contributing to the quality of spatial data within a GIS: imperfections (e.g., inaccuracy and imprecision) and effects of discretization. An example for discretization in the thematic domain is the chosen number of classes to represent a spatial phenomenon (e.g., air temperature). In order to improve the utility of a GIS an inclusion of a formal data quality model is essential. A data quality model stores, specifies, and handles the necessary data required to provide uncertainty information for GIS applications. This dissertation develops a data quality model that associates sources of uncertainty with units of information (e.g., measurement and coverage) in a GIS. The data quality model provides a basis to construct metrics

3 of 145 8.5 x 11 in

Start Microsoft PowerPoi... Adobe Acrobat - ... 11:09 AM

# Method2: make default tagged PDF and modify it to match ETD

The screenshot shows the Adobe Acrobat interface with the file 'Carpenter.pdf' open. The left sidebar contains a 'Tags' panel with a tree view of the document's structure. The main content area displays the 'Introduction' section, which is currently selected in the tree view. The text in the main area is rendered in a serif font and is partially obscured by a large black rectangular box, likely a redaction or a placeholder for a signature. The status bar at the bottom indicates the document is 10 of 60 pages, with a zoom level of 119% and a page size of 8.5 x 11 in.

**Tags Root**

- <front> Part2front
  - <title> P2title
  - <author> P2author
    - <given> new
      - Andrew
    - <surname> new
      - Carpenter
  - <abstract> new
    - Pulp and paper mill slud
- <body> Part2body
  - <chapter> Sect2chapter
    - <head> P2head
      - Introduction
    - <p> P2p
      - Treatment of the process
    - <p> P2p
      - Utilizing pulp sludge as
    - <li> LI2li
      - 1. organic matter, which
    - <LI>
    - <LI>
    - <P>
    - <P>
  - <Sect>
    - <H2>
    - <P>
    - <P>
    - <P>
    - <LI>

**Introduction**

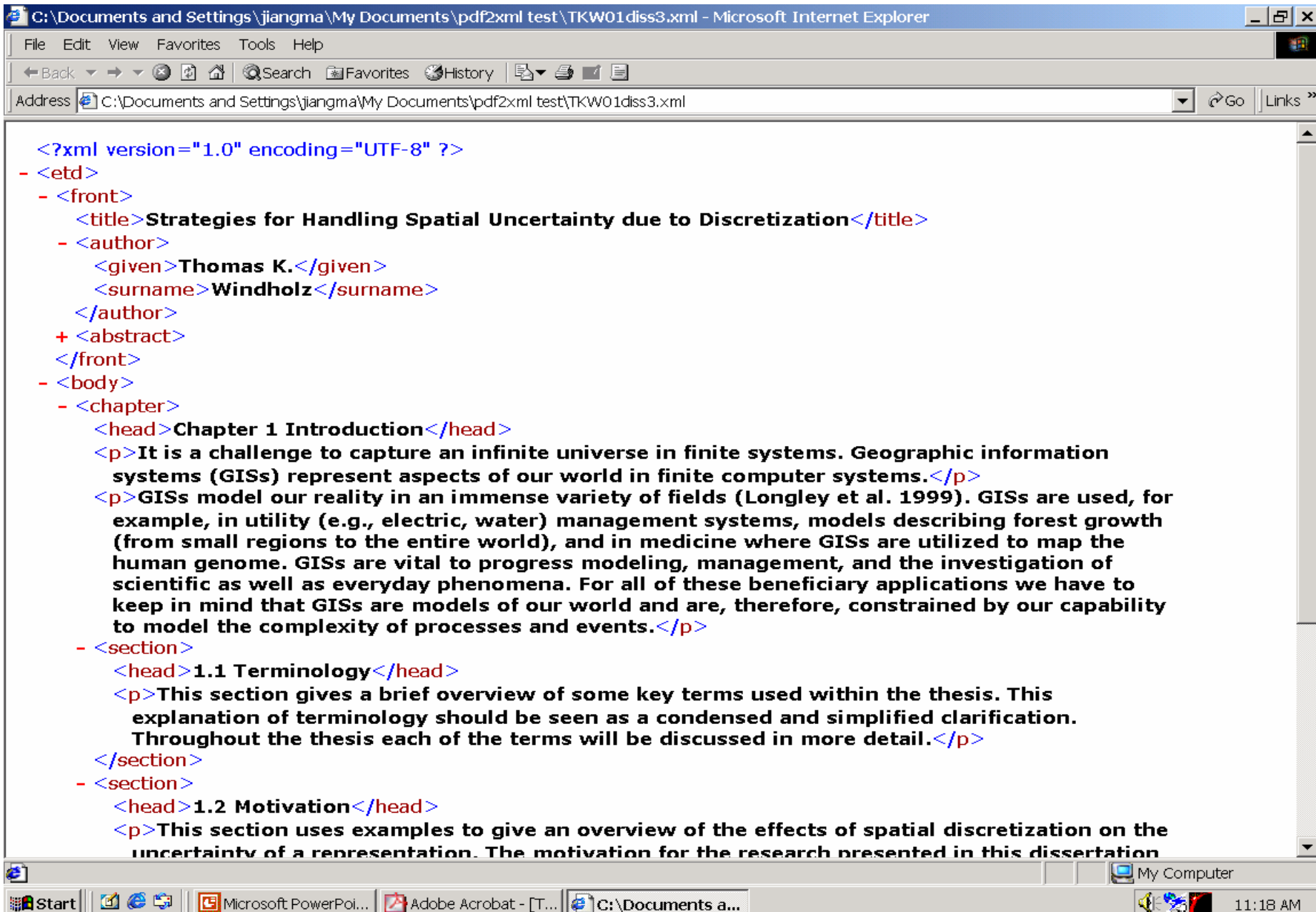
Treatment of the process water from the production of pulp generates large quantities of sludge which can represent a significant portion of the solid waste stream in paper-producing regions of the United States. In Maine alone, more than one million tons of pulp and paper mill sludge are

generated each year (Seekins, 1994). Pulp sludge consists primarily of wood fibers and the microbial biomass generated during the secondary wastewater treatment process. Pulp sludge is frequently disposed of in paper mill-owned or private landfills. Chemical analyses of pulp sludges indicate that concentrations of possible pollutants, such as heavy metals and organic contaminants, are quite low; consequently, paper mill managers are interested in exploring options for utilizing pulp sludge as a soil amendment. Pulp sludge is a good source of organic matter, which can improve the erosion-resistance and water-holding capacity of soils when used as a soil amendment. Recycling pulp sludge as a soil amendment may be an economically viable and environmentally sound long-term alternative to disposing of this material at a landfill.

Utilizing pulp sludge as a component of a manufactured topsoil is one such method of recycling. Topsoil manufacturing is the process of utilizing readily available materials to create a growing medium capable of sustaining desired vegetation. Compared to composting and other types of pulp sludge recycling, topsoil manufacturing requires less equipment and is less energy-intensive; consequently, it may be more cost-effective. The basic components of a manufactured topsoil consist of:

1. organic matter, which could be provided by the pulp sludge,
2. a mineral base material, which could be supplied by a common bank-run sand, and

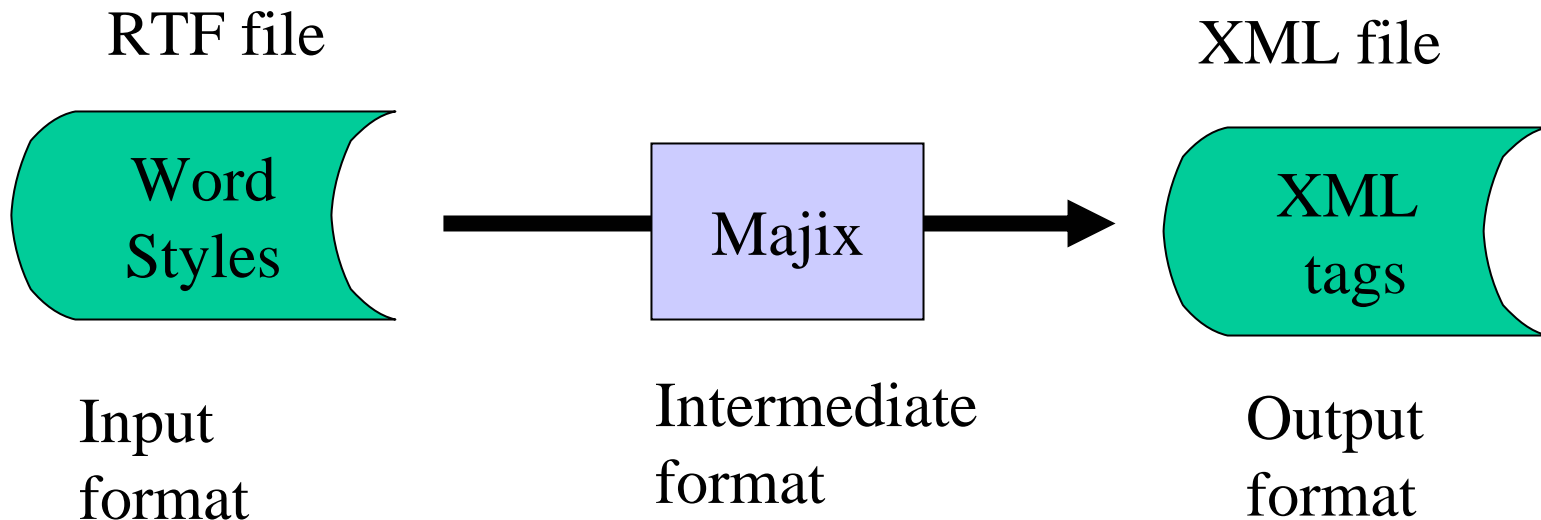
# The XML output after Save as XML



```
<?xml version="1.0" encoding="UTF-8" ?>
- <etd>
- <front>
  <title>Strategies for Handling Spatial Uncertainty due to Discretization</title>
  - <author>
    <given>Thomas K.</given>
    <surname>Windholz</surname>
  </author>
  + <abstract>
  </front>
- <body>
- <chapter>
  <head>Chapter 1 Introduction</head>
  <p>It is a challenge to capture an infinite universe in finite systems. Geographic information systems (GISs) represent aspects of our world in finite computer systems.</p>
  <p>GISs model our reality in an immense variety of fields (Longley et al. 1999). GISs are used, for example, in utility (e.g., electric, water) management systems, models describing forest growth (from small regions to the entire world), and in medicine where GISs are utilized to map the human genome. GISs are vital to progress modeling, management, and the investigation of scientific as well as everyday phenomena. For all of these beneficiary applications we have to keep in mind that GISs are models of our world and are, therefore, constrained by our capability to model the complexity of processes and events.</p>
- <section>
  <head>1.1 Terminology</head>
  <p>This section gives a brief overview of some key terms used within the thesis. This explanation of terminology should be seen as a condensed and simplified clarification. Throughout the thesis each of the terms will be discussed in more detail.</p>
</section>
- <section>
  <head>1.2 Motivation</head>
  <p>This section uses examples to give an overview of the effects of spatial discretization on the uncertainty of a representation. The motivation for the research presented in this dissertation
```

# Convert RTF to XML

## -- Majix



- can modify input style name
- can modify output xml tag name
- can not rearrange xml tree structure