CMSC 601
LaTeX 101
Spring 2011

Tim Finin
finin@cs.umbc.edu

Objective

• Understand the role of LaTeX in your research
• Learn how to create a simple LaTeX2e document
  — Create a LaTeX source file
  — Create and include figures
  — Reference figures and sections
  — Create lists
  — Include other tex files
  — Generate pdf output
  — Cite bibliographic references

History: TeX and LaTeX

• Donald Knuth created TeX in the late 70s so he could typeset his famous Art of Computer Programming books
• TeX produced great output and was very powerful (and programmable) but also very obscure
• Leslie Lamport of SRI produced LaTeX in the early 80s as a macro package making TeX easy to use
• I’ve never know anyone who used TeX directly

Other Options

• Microsoft Word is a great product
  — Track changes is a great feature
  — You can’t beat it for small documents
• HTML is fine if your target is a screen
  — The W3C does all of its documentation in HTML
  — The Kindle ebook format is HTML
• Google Docs is up and coming
  — great for real-time collaboration
• That’s about it these days
  — No one uses Tj6, Scribe, Pub, troff, WordPerfect, …
Why LaTeX?

• It’s good for complex documents like a dissertation
• It’s the standard for Computer Science, Mathematics and many other STEM fields
  — Many conferences have their own LaTeX document
  — Elsevier uses LaTeX to typeset all their journals
• LaTeX’s bibliography system, BibTeX, is the best
• LaTeX is programmable!
• LaTeX is open source software, has a large community of users and developers and a good infrastructure (e.g., CTAN)

Why LaTeX?

Accessing LaTeX

• LaTeX and associated tools are typically pre-installed on Linux and Mac OS X
• They are also on the CSEE servers and gl
• Use MikTeX for Windows

sample.tex

```latex
\documentclass[12pt]{article}
\usepackage{times}
\begin{document}
\title{Hello World in LaTeX}
\author{My Name Goes Here}
\maketitle
Hello, world!
\em Hello, world!
\bf Hello, world!
\Large \bf Hello, world!!!
\end{document}
```

Latex commands start with a backslash, required args are in {}, options in []

Start by declaring the document type (article) and use the 12pt option setting the font size

Loads required packages defining commands or setting parameters

LaTeX uses \begin{end} commands for blocks. Every document must have a document block

The title and author command set document variables and the \maketitle command generates the output text

Paragraphs are separated by blank lines

\{\} introduce un-named blocks and control scope. \em for italics, \bf for bold, \Large to increase font size
Compiling with pdflatex

> pdflatex sample
This is pdfTeX, Version 3.1415926-1.40.10 (TeX Live 2009)
entering extended mode
(/usr/local/texlive/2009/texmf-dist/tex/latex/base/article.cls
Document Class: article 2007/10/19 v1.4h Standard LaTeX
document class
... Output written on sample.pdf (1 page, 29675 bytes).
Transcript written on sample.log.

Compiling, old school

> latex sample
This is pdfTeX, Version 3.1415926-1.40.10 (TeX Live 2009)
Output written on sample.dvi (1 page, 652 bytes).
Transcript written on sample.log.
> dvips sample -o sample.ps
This is dvips(k) 5.98 Copyright 2009 Radical Eye Software
(www.radicaleye.com)
' TeX output 2011.01.31:0857' -> sample.ps
... > ps2pdf sample.ps

Output files

> Is -l sample*
-rw-r--r-- 1 finin staff 8 Jan 31 08:57 sample.aux
-rw-r--r-- 1 finin staff 652 Jan 31 08:57 sample.dvi
-rw-r--r-- 1 finin staff 3363 Jan 31 08:57 sample.log
-rw-r--r--@ 1 finin staff 3336 Jan 31 09:00 sample.pdf
-rw-r--r-- 1 finin staff 10664 Jan 31 08:58 sample.ps
-rw-r--r-- 1 finin staff 237 Jan 31 08:33 sample.tex

Files LaTeX Uses

• Input source file (.tex)
• Files containing structure and layout definitions (.sty)
• Tex formatted output file (.dvi)
• Others:
  .toc (table of contents), .lof (list of figures), .lot (list of
tables), .bib (bibliography)
Document Classes

- There are standard document classes: article, report, book, slides, letter
- Conferences and journals publish their own
  \documentclass{[10pt, journal, compsoc]{IEEEtran}
- These can be further customized via packages
  \usepackage{graphicx}
  \usepackage{algorithm}

Including Other LaTeX Files

- Supports modularity
  - a single LaTeX document can consist of multiple LaTeX files
  - Very useful for group work, e.g., many authors using SVN
- \input{intro}
  - used to include other LaTeX files
  - LaTeX filename is intro.tex

\begin{abstract}
\input{intro}
\end{abstract}

Real example

Vast amounts of information is available in structured forms like spreadsheets, database relations, and tables found in documents and on the Web. We describe an approach that uses linked data to interpret such tables and associate their components with nodes in a reference linked data collection. Our proposed framework assigns a class (i.e., type) to table columns, links table cells to entities, and inferred relations between columns to properties. The resulting interpretation can be used to annotate tables, confirm existing facts in the linked data collection, and propose new facts to be added. Our implemented prototype uses DBpedia as the linked data collection and Wikitology for existing facts in the linked data collection, and proposes new facts to be added. Our results are finding immediate applications in many areas, including improving information retrieval, text mining, and information extraction. Still more structured data is being extracted from text found on the web through several new research programs.

Real example
We did a preliminary evaluation for extraction of relation between columns. We asked human evaluators to identify pairs of columns in a table between which a relation may exist and compared that against the pairs of columns identified by the system. For five tables, used in this evaluation, in 25\% of the cases, the system was able to identify the correct pairs of columns.

## Conclusion

We presented an automated framework for interpreting data in a table using existing Linked Data KBs. Using the interpretation of the table we generate linked RDF from webtables. Evaluations show that we have been fairly successful in generating correct interpretation of webtables. Our current work is focused on improving relationship discovery and generating new facts and knowledge from tables that contain entities not present in the LOD knowledge bases. To deal with web scale analytics, we plan to focus on adapting our algorithms for parallelization using Hadoop or Azure type frameworks. We are also exploring ways to apply this work to create an automated (or semi-automated) human in the loop framework for interpreting and representing public government datasets as linked data.

## Bibliography

- ...