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?
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)
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
\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}
Compiling with pdflatex

> pdflatex sample
This is pdfTeX, Version 3.1415926-1.40.10 (TeX Live 2009)
exterting extended mode
./sample.tex
LaTeX2e <2009/09/24> ...
(/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.radicaleyeye.com)
'TeX output 2011.01.31:0857' -> sample.ps
...

> ps2pdf sample.ps
>
Output files

> ls -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
  \documentclass[11pt,letterpaper]{article}

• Conferences and journals publish their own
  \documentclass[10pt,journal,compsoc]{IEEEtran}
  \documentclass[runningheads,a4paper]{llncs}

• 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

A typical top level file

```latex
\documentclass[letterpaper]{article}
\usepackage{aaai}
\usepackage{times}
\usepackage{graphicx}
% comment: more here
\begin{document}
\include{title}
\include{intro}
\include{motivation}
\include{related}
\include{approach}
\include{evaluation}
\include{conclusion}
\include{bibliograph}
\end{document}
```
Real example

\documentclass[runningheads,a4paper]{llncs}
\usepackage{graphicx}
\usepackage{algorithm}
\usepackage{algorithmic}
\usepackage{subfigure}
\usepackage[rflt]{floatflt} \% floating figures
\usepackage{colortbl} \% for colors in tables
\setcounter{tocdepth}{3}
\begin{document}
\mainmatter
\title{Using linked data to interpret tables}\footnote{\scriptsize
Research supported in part by a gift from Microsoft Research, a
Fulbright fellowship, NSF award IIS-0326460 and the Human Language
Technology Center of Excellence.}}
\author{Varish Mulwad \and Tim Finin \and Zareen Syed \and Anupam Joshi}
\authorrunning{Varish Mulwad \and Tim Finin \and Zareen Syed \and Anupam Joshi}
\institute{Department of Computer Science and Electrical Engineering\\
University of Maryland, Baltimore County, Baltimore, MD USA 21250\\
\{varish1,finin,joshi\}@cs.umbc.edu, zareensyed@gmail.com}
\maketitle
Abstract
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 background knowledge. We evaluated its performance using a collection of tables from Google Squared, Wikipedia and the Web.

Section Introduction
Resources like Wikipedia and the Semantic Web's linked open data collection \cite{bizerc2009} are now being integrated to provide experimental knowledge bases containing both general purpose knowledge as well as a host of specific facts about significant people, places, organizations, events and many other entities of interest. The 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 \cite{etzioni2006machine,mcnamee2009overview}.
of the class labels predicted were considered correct by the evaluators. The accuracy in each of the four categories is shown in Figure \ref{columnCorrectness_EntityLinking}. We enjoyed moderate success in assigning class labels for \textit{Organizations} and \textit{Other} types of data probably because of sparseness of data in the KB about these types of entities.

\begin{figure}[tbp]
\fbox{\includegraphics[scale = 0.65]{images/accuracy_chart}}
\caption{Category wise accuracy for ``column correctness'' is shown in (a) and for entity linking in (b) }
\label{columnCorrectness_EntityLinking}
\end{figure}

\subsection{Linking table cells to entities}

For the evaluation of linking table cells to entities, we manually hand-labeled the 611 table cells to their appropriate Wikipedia / DBpedia pages. The system generated links were compared against the
\subsection{Relation identification}

We did a preliminary evaluation for identification 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.

\section{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.

\bibliographystyle{springer}
\bibliography{cold}
@article{bizer2009,
  author = {Bizer, Christian},
  journal = {IEEE Intelligent Systems},
  number = {5},
  pages = {87--92},
  title = {The Emerging Web of Linked Data},
  volume = {24},
  year = {2009} }

@inproceedings{zieglerp04,
  author = {Ziegler, Patrick and Dittrich, Klaus R.},
  booktitle = {Building the Information Society},
  doi = {10.1007/978-1-4020-8157-6_1},
  pages = {3--12},
  publisher = {Springer Boston},
  title = {Three Decades of Data Integration: all Problems Solved?},
  url = {http://www.springerlink.com/content/t25x6t660v43m37k/},
  volume = {156},
  year = {2004} }

@MastersThesis{t2ldvishthesis10,
  author = "Varish Mulwad",
  title = {{T2LD} - An automatic framework for extracting, interpreting and representing tables as Linked Data},
  month = "August",
  year = "2010",
  publisher = "UMBC",
  school= {{U. of Maryland, Baltimore County}},

Real example
cold.bib
LaTeX Miscellanea

• Latex quotes and hyphens
  – ``'' – *not* " "
  – --- or ---, *not* –

• Capitalizing words in BibTeX references
  – Use {} to force capitalization
  – title = "A New Algorithm for Bayesian MCMC Inference"
    • A new algorithm for bayesian mcmc inference
  – title = "A New Algorithm for \{Bayesian MCMC\} Inference"
    • A new algorithm for Bayesian MCMC inference