COMP90055 Computing Project Report

SurName: Teng

Given Name: Ruichen

Student Number: 678693

University Email: tengr@student.unimelb.edu.au

Name of Degree Enrolled in: Master of Information Technology — Computing

Subject Code: COMP90055 Computing Project

Total credit points for entire project: 25

Type of Project: Research project

Semester in which project commenced: Semester 1, 2015

Semester in which project is expected to complete: Semester1, 2015

Project Title: Information Extraction of biomedical relationships in published colon cancer literature

Supervisor: Prof. Karin Verspoor

Information Extraction of biomedical relationships in published colon cancer literature

Ruichen Teng

Department of Computing and Information Systems

University of Melbourne

This dissertation is submitted for the degree of

Master of Information Technology

June 2015

To my loving parents . . .

Declaration

I certify that

- this thesis does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person where due reference is not made in the text.

- where necessary I have received clearance for this research from the University's Ethics Committee (Approval Number ....) and have submitted all required data to the Department

- the thesis is 8000 words in length (excluding text in images, table, bibliographies and appendices).

Ruichen Teng
June 2015

Acknowledgements

And I would like to acknowledge ...

Abstract

This is where you write your abstract ...

Table of contents

List of figures

List of tables

Nomenclature

Roman Symbols

F	complex function
F	complex function

Greek Symbols

γ	a simply closed curve on a complex plane
γ	a simply closed curve on a complex plane
$\imath$	unit imaginary number $\sqrt{-1}$
$\imath$	unit imaginary number $\sqrt{-1}$
π	$\simeq 3.14\ldots$
π	$\simeq 3.14\ldots$

Superscripts

j	superscript index
j	superscript index

Subscripts

0	subscript index
0	subscript index

Other Symbols

$\oint_\gamma$	integration around a curve γ

$\oint_\gamma$ integration around a curve γ

Acronyms / Abbreviations

CIF Cauchy's Integral Formula

CIF Cauchy's Integral Formula

Introduction

1.1 Motivation

Text mining is the process of searching for patterns in natural language text using methods in computer science, linguistics, and statistics. Despite being unstructured and only human-understandable, text is still our primary media for exchange of information[13]. The prevalence of textual data presents a big challenge to computer-driven natural language understanding. *Information extraction*, in particular, refers to the task of acquiring organized, structured and queryable format of data from the unstructured corpus.

While text mining is widely used in areas like marketing and document verifying, it has received increased attention for its application to biomedical literatures[1, 5, 6]. This trend stems from the direct need of biomedical workers and researchers to cope with information explosion in their field. For instance, MEDLINE(Medical Literature Analysis andRetrieval System Online), the online database of United States National Library of Medicine, has accumulated nearly 0.8 million citations and 2.7 billion searches in 2014 alone[7], with total citations reaching 22 million. Within these publications there are valuable research results that should add to human knowledge. In the meantime, our primary knowledge base in life science - the biomedical databases, are still mostly being populated manually by *biocurators* - the "museum catalogers of the Internet age"[12]. They are professional scientists who read biomedical articles, record relevant data and organize them according to the biomedical database schema. The sheer volume of publications has made this process increasingly unrealistic[4].

Not only does data overload make knowledge discovery demanding, it also leads to a decline in literature quality. Nowadays biomedical workers and researchers are more prone to drawing wrong conclusions because they simply can not read all the relevant publications, among which oftentimes contradicting results are reported. Needless to say, we are

in desperate need of automatic tools for systematically analyzing documents and extracting information. In fact, it has been argued that text mining is required to improve the coverage of databases[3].

1.2 Definitions and Assumptions

Below are a list of terms which I will use throughout this thesis, detailed examples will be given in the relevant sections, but a general definition is first given here to avoid any confusions that might arise when reading the following paragraphs.

- *Relation Extraction:* In general, relation extraction refers two the process of discovering a relationship between entities in text. In the domain of natural language processing, the relation can be semantic, syntactic, etc, with semantic relations being the most important for knowledge discovery. Relations can be unary, binary and complex with complex relations sometimes intermingled with the concept of events, In this project we are mainly concerned with binary relations as shown. Relation extraction is a form of information extraction where the semantic relations between entities are extracted. Specifically, in this project we focus on the relation extraction among other informations. Biomedical relations covers a wide range of knowledge in this field.

- *Event Extraction:* to be added

1.3 Research Question

This project looks to investigate the application of information extraction system Approximate Subgraph Matching (ASM)[**?**], on relation extraction tasks, specifically regarding the relation extraction on the variome corpus. In this project we focus on the relation extraction among other informations. Biomedical relations covers a wide range of knowledge in this field. As it turns out, this is not a trivial process, The relation extraction tool has a very promising applications for researchers and medical field workers, pharmaceutical companies, and the general public.

1.4 Thesis Structure

The thesis is organized as follows: chapter 1 will

therefore to allow researchers to identify needed information more efficiently, quote Based on the information in 1.1 and 1.2, the aim of this project is to develop tools that can assist the human bio-curation process.

Related Work

This chapter will explore the research related to this thesis.

2.1 Relation Extraction

In general, relation extraction refers two the process of discovering a relationship between entities in text. In the domain of natural language processing, the relation can be semantic, syntactic, etc, with semantic relations being the most important for knowledge discovery. Relations can be uniary, binary and complext with complex relations sometimes intermingled with the concept of events, In this project we are mainly concerned with binary relations as shown

2.2 Named entity recognition

Named Entity Recognition, or NER, is the task of identifying elements in text that belong to pre-defined categories like person, organization, etc. Specifically, NER in biomedical text mining aims at identifying thing like proteins, diseases, genes, etc. There is extra difficulty for NER in the biomeidical domain mainly because of the following reasons.

quote: A survey of current work in biomedical text mining This task has been challenging for several reasons. First, there does not exist a complete dictionary for most types of biological named entities, so simple text- matching algorithms do not suffice. In addition, the same word or phrase can Recognising biological entities in text allows for further extraction of relationships and other information by identifying the key concepts of interest refer to a different thing depending upon context (eg ferritin can be a biological substance or a laboratory test). Conversely, many biological entities have several names (eg PTEN and MMAC1 refer the same gene). Biological entities may also have multi-word names (eg carotid artery), so the problem is additionally complicated by the need to determine name boundaries and resolve overlap of candidate names. Because of the potential utility and

complexity of the problem, NER has attracted the interest of many researchers, and there is a tremendous amount of published research in this topic. With the large amount of genomic information being generated by biomedical researchers, it should not be surprising that in the genomics era, much of the work in biomedical NER has focused on recognising gene and protein names in free text. quote: A survey of current work in biomedical text mining

2.3 Pattern based Methods

Pattern based methods

2.4 Co-corrence based methods

2.5 Rule-Based Methods

2.6 Kernel based methods

2.7 Feature based methods

2.8

Core

3.1 Data Collection

Our dataset is the Variome Corpus[11], which is openly accessible. [1] Verspoor et al. [11] gave a detailed illustration of the document selection and annotation process. I will summarize the main points here.

3.1.1 Background

A major part of the current biomedical research lies in understanding the relations between human genetic variation and disease phenotypes. The *Human Variome Project*, or *HVP*, is a global initiative to collect all genetic variation information affecting human health[9]. In particular, it acts as a liaison between individuals and organizations to integrate the genetic variants into databases that are open to the general public[11]. The *International Society for Gastrointestinal Hereditary Tumours (InSiGHT)*, is an international organization which aims to benefit patients with hereditary gastrointestinal(GI) tumours by research, education and personal assistance. In 2008, InSiGHT and HVP began a collaboration which propels InSiGHT to refine its process in the integration and interpretation of genetic variants. Consequently, a substantial effort was made to understand the mutation of mismatch repair(MMR) genes, the cause of Lynch Syndrome - one of the main syndromes of GI cancer[10]. A total of 10 full-text articles were selected from PubMed Central®by searching the common Lynch syndrome genes. These documents are mostly about inherited colon cancer. The annotation schema, also known as the Variome Annotation Schema[11], include 11 entity types and 13 relation types.

In short, the corpus is Inspired by needs of inSIGHT database, but intended for broader applications Documents relevant to the genetics of Lynch syndrome, which covers inherited colon cancer as well as certain other cancers.Selected with PubMed Central To train tools for

[1] http://www.opennicta.com.au/home/health/variome

mining genetic variation and its relationship to disease Here in this information extraction task, we treat the manually annotated data as the *gold data*,

3.2 Algorithm

Named entity recoginition is the process of Iis project cite the performance of the subgraph matching method, as an instance-based learning strategy (Alpaydin, 2004), is dependent on having good training examples that express the events in a range of syntactic structures, cite

Related Work

4.1 Relation Extraction

cite The focus must be more on helping biomedical researchers to solve real-world problems that are inhibiting the pace of research and less on evaluations based on system output independent of meeting user needs. cite

4.2 Definitions and Assumptions

4.3 Named entity recognition

Named entity recognition is the process of Iis project

4.4 Pattern based Methods

4.5 Co-corrence based methods

Syntactic parsing is particularly useful for relations that are distant in their expression. cite Distant supervision for relation extraction without labeled data cite

4.6 Kernel based methods

4.7 Feature based methods

4.8

Getting started

5.1 Corpus Construction

The corpus consists of 10 full text publicaitons, known as the Variome Corpus[11]. The Variome copurs is created partly inline with the database schema of the Internation Society for Gasterinjoal Hereidateyy Databse, but with the posibility of broader applications in mind. In total there are 11 entity types and 13 relation types, as shown.

5.2 System Adaptation

5.3 Where does it come from?

Contrary to popular belief, Lorem Ipsum is not simply random text. It has roots in a piece of classical Latin literature from 45 BC, making it over 2000 years old. Richard McClintock, a Latin professor at Hampden-Sydney College in Virginia, looked up one of the more obscure Latin words, consectetur, from a Lorem Ipsum passage, and going through the cites of the word in classical literature, discovered the undoubtable source. Lorem Ipsum comes from sections 1.10.32 and 1.10.33 of "de Finibus Bonorum et Malorum" (The Extremes of Good and Evil) by Cicero, written in 45 BC. This book is a treatise on the theory of ethics, very popular during the Renaissance. The first line of Lorem Ipsum, "Lorem ipsum dolor sit amet..", comes from a line in section 1.10.32.

The standard chunk of Lorem Ipsum used since the 1500s is reproduced below for those interested. Sections 1.10.32 and 1.10.33 from “de Finibus Bonorum et Malorum" by Cicero are also reproduced in their exact original form, accompanied by English versions from the 1914 translation by H. Rackham

“Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation

ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum."

Section 1.10.32 of “de Finibus Bonorum et Malorum", written by Cicero in 45 BC: “Sed ut perspiciatis unde omnis iste natus error sit voluptatem accusantium doloremque laudantium, totam rem aperiam, eaque ipsa quae ab illo inventore veritatis et quasi architecto beatae vitae dicta sunt explicabo. Nemo enim ipsam voluptatem quia voluptas sit aspernatur aut odit aut fugit, sed quia consequuntur magni dolores eos qui ratione voluptatem sequi nesciunt. Neque porro quisquam est, qui dolorem ipsum quia dolor sit amet, consectetur, adipisci velit, sed quia non numquam eius modi tempora incidunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim ad minima veniam, quis nostrum exercitationem ullam corporis suscipit laboriosam, nisi ut aliquid ex ea commodi consequatur? Quis autem vel eum iure reprehenderit qui in ea voluptate velit esse quam nihil molestiae consequatur, vel illum qui dolorem eum fugiat quo voluptas nulla pariatur?"

1914 translation by H. Rackham: “But I must explain to you how all this mistaken idea of denouncing pleasure and praising pain was born and I will give you a complete account of the system, and expound the actual teachings of the great explorer of the truth, the master-builder of human happiness. No one rejects, dislikes, or avoids pleasure itself, because it is pleasure, but because those who do not know how to pursue pleasure rationally encounter consequences that are extremely painful. Nor again is there anyone who loves or pursues or desires to obtain pain of itself, because it is pain, but because occasionally circumstances occur in which toil and pain can procure him some great pleasure. To take a trivial example, which of us ever undertakes laborious physical exercise, except to obtain some advantage from it? But who has any right to find fault with a man who chooses to enjoy a pleasure that has no annoying consequences, or one who avoids a pain that produces no resultant pleasure?"

Section 1.10.33 of “de Finibus Bonorum et Malorum", written by Cicero in 45 BC: “At vero eos et accusamus et iusto odio dignissimos ducimus qui blanditiis praesentium voluptatum deleniti atque corrupti quos dolores et quas molestias excepturi sint occaecati cupiditate non provident, similique sunt in culpa qui officia deserunt mollitia animi, id est laborum et dolorum fuga. Et harum quidem rerum facilis est et expedita distinctio. Nam libero tempore, cum soluta nobis est eligendi optio cumque nihil impedit quo minus id quod maxime placeat facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et aut officiis debitis aut rerum necessitatibus saepe eveniet ut et voluptates repudiandae sint et molestiae non recusandae. Itaque earum rerum hic tenetur a

sapiente delectus, ut aut reiciendis voluptatibus maiores alias consequatur aut perferendis doloribus asperiores repellat."

1914 translation by H. Rackham: “On the other hand, we denounce with righteous indignation and dislike men who are so beguiled and demoralized by the charms of pleasure of the moment, so blinded by desire, that they cannot foresee the pain and trouble that are bound to ensue; and equal blame belongs to those who fail in their duty through weakness of will, which is the same as saying through shrinking from toil and pain. These cases are perfectly simple and easy to distinguish. In a free hour, when our power of choice is untrammelled and when nothing prevents our being able to do what we like best, every pleasure is to be welcomed and every pain avoided. But in certain circumstances and owing to the claims of duty or the obligations of business it will frequently occur that pleasures have to be repudiated and annoyances accepted. The wise man therefore always holds in these matters to this principle of selection: he rejects pleasures to secure other greater pleasures, or else he endures pains to avoid worse pains."

My second chapter

6.1 Reasonably long section title

I'm going to randomly include a picture Figure 6.1.

If you have trouble viewing this document contact Krishna at: kks32@cam.ac.uk or raise an issue at https://github.com/kks32/phd-thesis-template/

Enumeration

1. The first topic is dull
2. The second topic is duller
    (a) The first subtopic is silly
    (b) The second subtopic is stupid
3. The third topic is the dullest

itemize

- The first topic is dull
- The second topic is duller
    - The first subtopic is silly
    - The second subtopic is stupid
- The third topic is the dullest

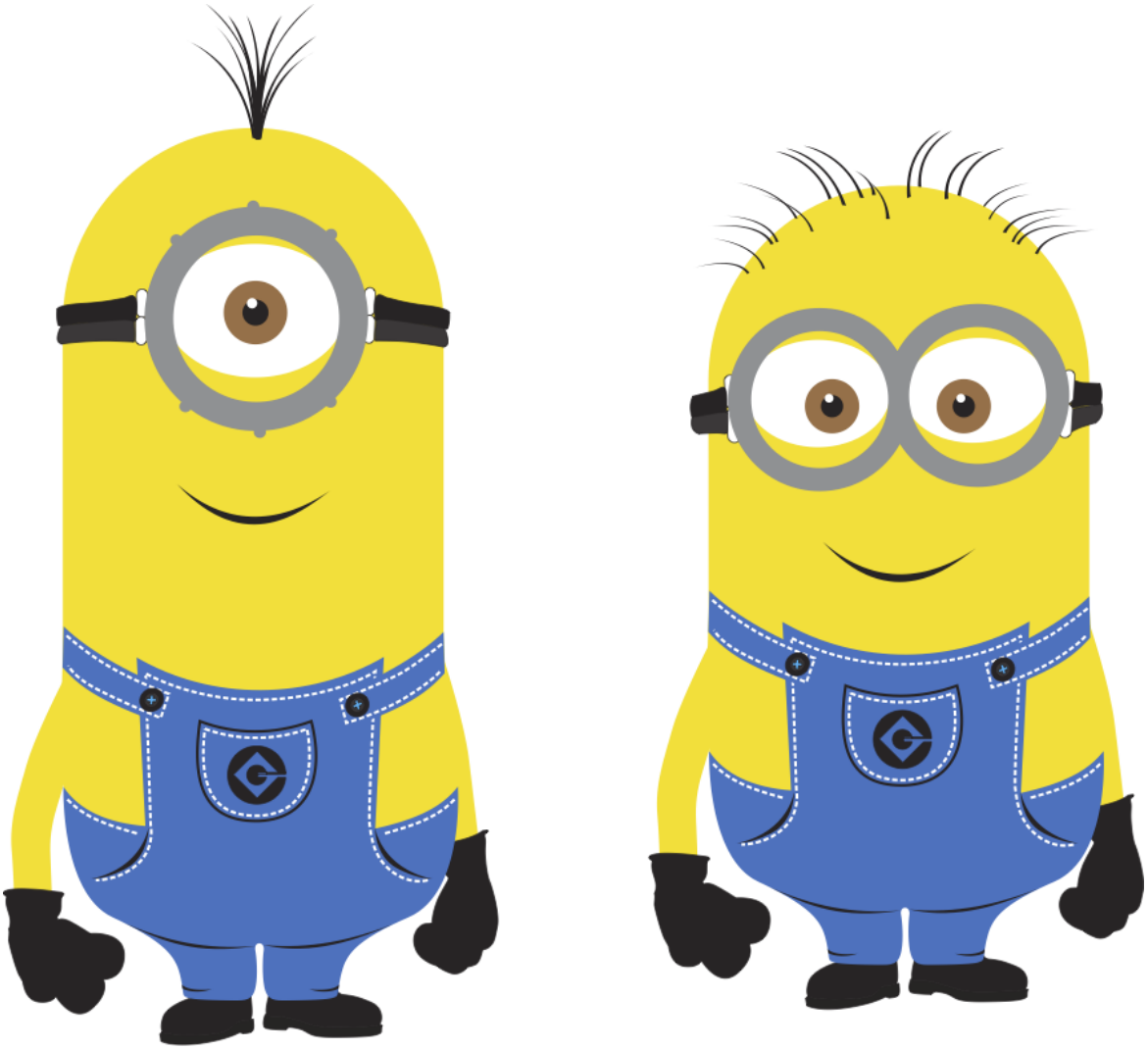


Fig. 6.1 This is just a long figure caption for the minion in Despicable Me from Pixar

description

The first topic is dull

The second topic is duller

- **The first subtopic** is silly
- **The second subtopic** is stupid

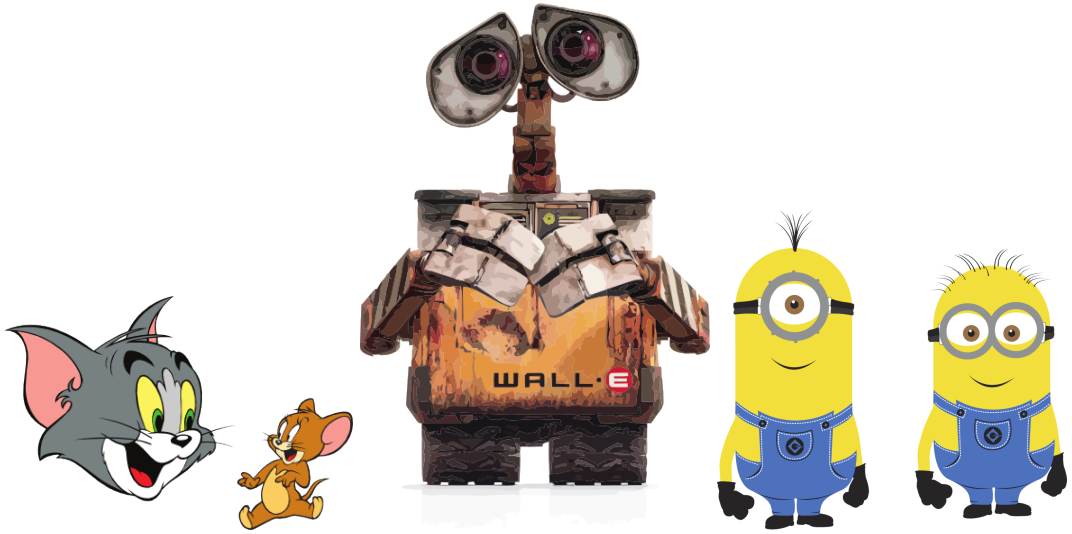
The third topic is the dullest

6.2 Hidden section

Lorem ipsum dolor sit amet, *consectetur adipiscing elit*. In magna nisi, aliquam id blandit id, congue ac est. Fusce porta consequat leo. Proin feugiat at felis vel consectetur. Ut tempus ipsum sit amet congue posuere. Nulla varius rutrum quam. Donec sed purus luctus, faucibus velit id, ultrices sapien. Cras diam purus, tincidunt eget tristique ut, egestas quis nulla. Curabitur vel iaculis lectus. Nunc nulla urna, ultrices et eleifend in, accumsan ut erat. In ut ante leo. Aenean a lacinia nisl, sit amet ullamcorper dolor. Maecenas blandit, tortor ut scelerisque congue, velit diam volutpat metus, sed vestibulum eros justo ut nulla. Etiam nec ipsum non enim luctus porta in in massa. Cras arcu urna, malesuada ut tellus ut, pellentesque mollis risus.Morbi vel tortor imperdiet arcu auctor mattis sit amet eu nisi. Nulla gravida urna vel nisl egestas varius. Aliquam posuere ante quis malesuada dignissim. Mauris ultrices tristique eros, a dignissim nisl iaculis nec. Praesent dapibus tincidunt mauris nec tempor. Curabitur et consequat nisi. Quisque viverra egestas risus, ut sodales enim blandit at. Mauris quis odio nulla. Cras euismod turpis magna, in facilisis diam congue non. Mauris faucibus nisl a orci dictum, et tempus mi cursus.

Etiam elementum tristique lacus, sit amet eleifend nibh eleifend sed [1]. Maecenas dapibu augue ut urna malesuada, non tempor nibh mollis. Donec sed sem sollicitudin, convallis velit aliquam, tincidunt diam. In eu venenatis lorem. Aliquam non augue porttitor tellus faucibus porta et nec ante. Proin sodales, libero vitae commodo sodales, dolor nisi cursus magna, non tincidunt ipsum nibh eget purus. Nam rutrum tincidunt arcu, tincidunt vulputate mi sagittis id. Proin et nisi nec orci tincidunt auctor et porta elit. Praesent eu dolor ac magna cursus euismod. Integer non dictum nunc.

[1] My footnote goes blah blah blah! ...



(a) Tom and Jerry (b) Wall-E (c) Minions

Fig. 6.2 Best Animations

Subplots

I can cite Wall-E (see Fig. 6.2b) and Minions in despicable me (Fig. 6.2c) or I can cite the whole figure as Fig. 6.2

My third chapter

7.1 First section of the third chapter

And now I begin my third chapter here . . .

And now to cite some more people Ancey et al. [2], Read [8]

7.1.1 First subsection in the first section

. . . and some more

7.1.2 Second subsection in the first section

. . . and some more . . .

First subsub section in the second subsection

. . . and some more in the first subsub section otherwise it all looks the same doesn't it? well we can add some text to it . . .

7.1.3 Third subsection in the first section

. . . and some more . . .

First subsub section in the third subsection

. . . and some more in the first subsub section otherwise it all looks the same doesn't it? well we can add some text to it and some more and some more and some more and some more and some more and some more and some more . . .

Second subsub section in the third subsection

. . . and some more in the first subsub section otherwise it all looks the same doesn't it? well we can add some text to it . . .

7.2 Second section of the third chapter

and here I write more . . .

7.3 The layout of formal tables

This section has been modified from "Publication quality tables in LaTeX*" by Simon Fear.

The layout of a table has been established over centuries of experience and should only be altered in extraordinary circumstances.

When formatting a table, remember two simple guidelines at all times:

1. Never, ever use vertical rules (lines).
2. Never use double rules.

These guidelines may seem extreme but I have never found a good argument in favour of breaking them. For example, if you feel that the information in the left half of a table is so different from that on the right that it needs to be separated by a vertical line, then you should use two tables instead. Not everyone follows the second guideline:

There are three further guidelines worth mentioning here as they are generally not known outside the circle of professional typesetters and subeditors:

3. Put the units in the column heading (not in the body of the table).
4. Always precede a decimal point by a digit; thus 0.1 *not* just .1.
5. Do not use 'ditto' signs or any other such convention to repeat a previous value. In many circumstances a blank will serve just as well. If it won't, then repeat the value.

A frequently seen mistake is to use '\begin{center}' . . . '\end{center}' inside a figure or table environment. This center environment can cause additional vertical space. If you want to avoid that just use '\centering'

Table 7.1 A badly formatted table

	Species I		Species II	
Dental measurement	mean	SD	mean	SD
I1MD	6.23	0.91	5.2	0.7
I1LL	7.48	0.56	8.7	0.71
I2MD	3.99	0.63	4.22	0.54
I2LL	6.81	0.02	6.66	0.01
CMD	13.47	0.09	10.55	0.05
CBL	11.88	0.05	13.11	0.04

Table 7.2 A nice looking table

Dental measurement	Species I		Species II	
	mean	SD	mean	SD
I1MD	6.23	0.91	5.2	0.7
I1LL	7.48	0.56	8.7	0.71
I2MD	3.99	0.63	4.22	0.54
I2LL	6.81	0.02	6.66	0.01
CMD	13.47	0.09	10.55	0.05
CBL	11.88	0.05	13.11	0.04

Table 7.3 Even better looking table using booktabs

Dental measurement	Species I		Species II	
	mean	SD	mean	SD
I1MD	6.23	0.91	5.2	0.7
I1LL	7.48	0.56	8.7	0.71
I2MD	3.99	0.63	4.22	0.54
I2LL	6.81	0.02	6.66	0.01
CMD	13.47	0.09	10.55	0.05
CBL	11.88	0.05	13.11	0.04

References

[1] Ananiadou, S., Kell, D. B., and Tsujii, J.-i. (2006). Text mining and its potential applications in systems biology. *Trends in biotechnology*, 24(12):571–579.

[2] Ancey, C., Coussot, P., and Evesque, P. (1996). Examination of the possibility of a fluid-mechanics treatment of dense granular flows. *Mechanics of Cohesive-frictional Materials*, 1(4):385–403.

[3] Baumgartner, W. A., Cohen, K. B., Fox, L. M., Acquaah-Mensah, G., and Hunter, L. (2007). Manual curation is not sufficient for annotation of genomic databases. *Bioinformatics*, 23(13):i41–i48.

[4] Cohen, A. M. and Hersh, W. R. (2005). A survey of current work in biomedical text mining. *Briefings in bioinformatics*, 6(1):57–71.

[5] Kim, J.-D., Ohta, T., Tateisi, Y., and Tsujii, J. (2003). Genia corpus—a semantically annotated corpus for bio-textmining. *Bioinformatics*, 19(suppl 1):i180–i182.

[6] Krallinger, M. and Valencia, A. (2005). Text-mining and information-retrieval services for molecular biology. *Genome biology*, 6(7):224.

[7] MEDLINE (2015). Key medline® indicators.

[8] Read, C. J. (1985). A solution to the invariant subspace problem on the space l_1. *Bull. London Math. Soc.*, 17:305–317.

[9] Ring, H. Z., Kwok, P.-Y., and Cotton, R. (2006). Human variome project: an international collaboration to catalogue human genetic variation. *Pharmacogenomics*, 7(7):969–972.

[10] Silva, F. C. C. d., Valentin, M. D., Ferreira, F. d. O., Carraro, D. M., and Rossi, B. M. (2009). Mismatch repair genes in lynch syndrome: a review. *Sao Paulo Medical Journal*, 127(1):46–51.

[11] Verspoor, K., Yepes, A. J., Cavedon, L., McIntosh, T., Herten-Crabb, A., Thomas, Z., and Plazzer, J.-P. (2013). Annotating the biomedical literature for the human variome. *Database*, 2013:bat019.

[12] Wikipedia (2014). Biocurator — wikipedia, the free encyclopedia. [Online; accessed 31-May-2015].

[13] Witten, I. H. (2005). Text mining. *Practical handbook of Internet computing*, pages 14–1.

How to install LaTeX

Windows OS

TeXLive package - full version

1. Download the TeXLive ISO (2.2GB) from
   https://www.tug.org/texlive/
2. Download WinCDEmu (if you don't have a virtual drive) from
   http://wincdemu.sysprogs.org/download/
3. To install Windows CD Emulator follow the instructions at
   http://wincdemu.sysprogs.org/tutorials/install/
4. Right click the iso and mount it using the WinCDEmu as shown in
   http://wincdemu.sysprogs.org/tutorials/mount/
5. Open your virtual drive and run setup.pl

or

Basic MikTeX - TeX distribution

1. Download Basic-MiKTeX(32bit or 64bit) from
   http://miktex.org/download
2. Run the installer
3. To add a new package go to Start » All Programs » MikTex » Maintenance (Admin)
   and choose Package Manager
4. Select or search for packages to install

TexStudio - TeX editor

1. Download TexStudio from
   http://texstudio.sourceforge.net/#downloads

2. Run the installer

Mac OS X

MacTeX - TeX distribution

1. Download the file from
   https://www.tug.org/mactex/

2. Extract and double click to run the installer. It does the entire configuration, sit back and relax.

TexStudio - TeX editor

1. Download TexStudio from
   http://texstudio.sourceforge.net/#downloads

2. Extract and Start

Unix/Linux

TeXLive - TeX distribution

Getting the distribution:

1. TexLive can be downloaded from
   http://www.tug.org/texlive/acquire-netinstall.html.

2. TexLive is provided by most operating system you can use (rpm,apt-get or yum) to get TexLive distributions

Installation

1. Mount the ISO file in the mnt directory

   ```
   mount -t iso9660 -o ro,loop,noauto /your/texlive####.iso /mnt
   ```

2. Install wget on your OS (use rpm, apt-get or yum install)

3. Run the installer script install-tl.

   ```
   cd /your/download/directory
   ./install-tl
   ```

4. Enter command 'i' for installation

5. Post-Installation configuration:
   http://www.tug.org/texlive/doc/texlive-en/texlive-en.html#x1-320003.4.1

6. Set the path for the directory of TexLive binaries in your .bashrc file

For 32bit OS

For Bourne-compatible shells such as bash, and using Intel x86 GNU/Linux and a default directory setup as an example, the file to edit might be

```
edit $~/.bashrc file and add following lines
PATH=/usr/local/texlive/2011/bin/i386-linux:$PATH;
export PATH
MANPATH=/usr/local/texlive/2011/texmf/doc/man:$MANPATH;
export MANPATH
INFOPATH=/usr/local/texlive/2011/texmf/doc/info:$INFOPATH;
export INFOPATH
```

For 64bit OS

```
edit $~/.bashrc file and add following lines
PATH=/usr/local/texlive/2011/bin/x86_64-linux:$PATH;
export PATH
MANPATH=/usr/local/texlive/2011/texmf/doc/man:$MANPATH;
export MANPATH
```

```
INFOPATH=/usr/local/texlive/2011/texmf/doc/info:$INFOPATH;
export INFOPATH
```

Fedora/RedHat/CentOS:

```
sudo yum install texlive
sudo yum install psutils
```

SUSE:

```
sudo zypper install texlive
```

Debian/Ubuntu:

```
sudo apt-get install texlive texlive-latex-extra
sudo apt-get install psutils
```

Installing the CUED class file

LaTeX.cls files can be accessed system-wide when they are placed in the <texmf>/tex/latex directory, where <texmf> is the root directory of the user's TeXinstallation. On systems that have a local texmf tree (<texmflocal>), which may be named "texmf-local" or "localtexmf", it may be advisable to install packages in <texmflocal>, rather than <texmf> as the contents of the former, unlike that of the latter, are preserved after the LaTeXsystem is reinstalled and/or upgraded.

It is recommended that the user create a subdirectory <texmf>/tex/latex/CUED for all CUED related LaTeXclass and package files. On some LaTeXsystems, the directory look-up tables will need to be refreshed after making additions or deletions to the system files. For TeXLive systems this is accomplished via executing "texhash" as root. MIKTeXusers can run "initexmf -u" to accomplish the same thing.

Users not willing or able to install the files system-wide can install them in their personal directories, but will then have to provide the path (full or relative) in addition to the filename when referring to them in LaTeX.

Index