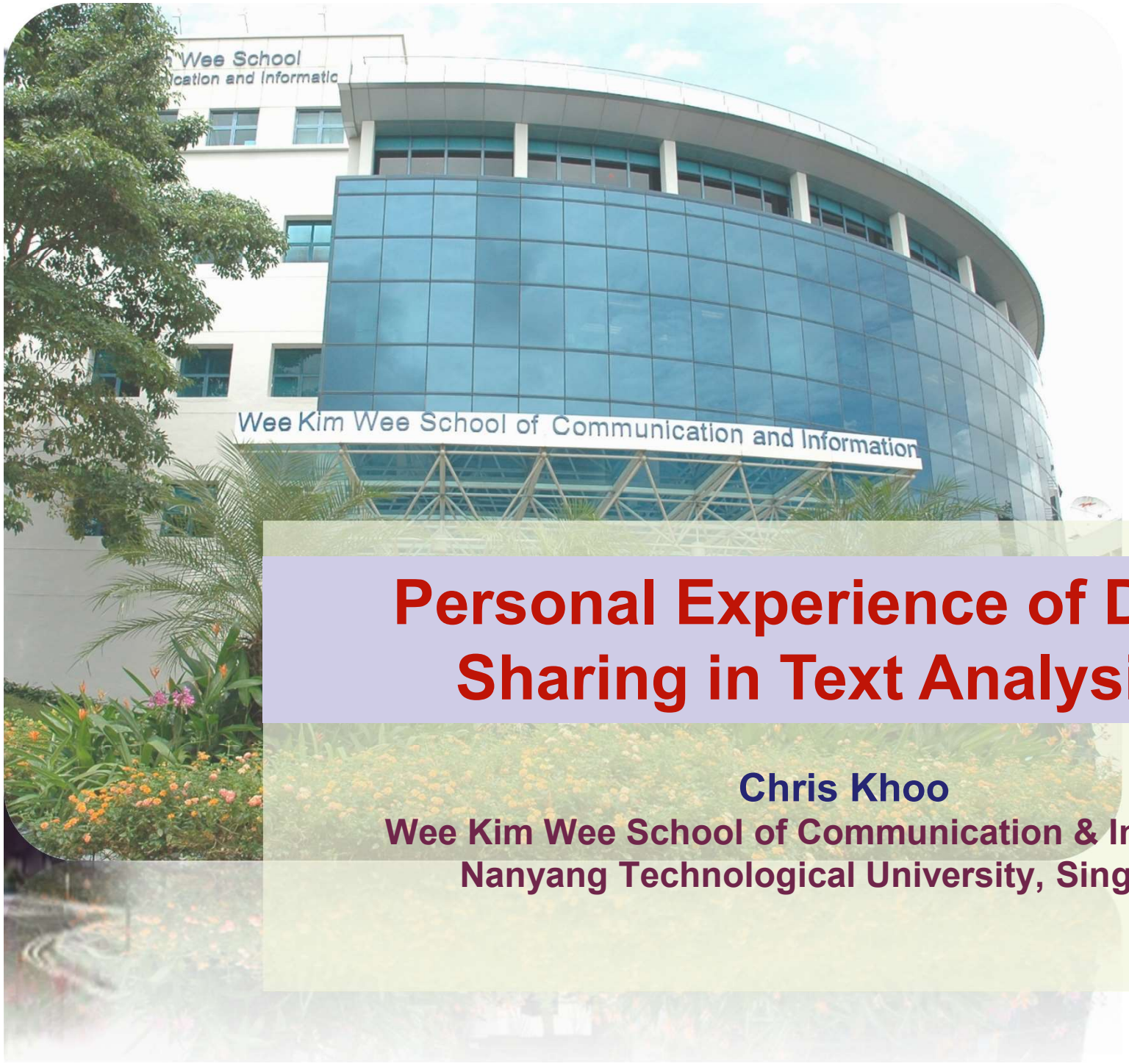


Personal Experience of Data Sharing in Text Analysis

Chris Khoo
Wee Kim Wee School of Communication & Information
Nanyang Technological University, Singapore

Why share data?

1. **Retirement is looming**
   - What impact have I had on my field, except for a few papers?
2. **NTU is #1 in Asia**
   - The best universities share resources with the research community
   - a way of providing leadership and supporting progress in the research community
3. **Share resources with public to support educational use** (and recreational and industry use)
4. **Support intelligent Web**
   - Semantic web and linked data applications

Resources in text mining and natural language processing

- Lexical resources: lists of words/phrases with semantic categorization
  - Sentiment lexicon
  - List of causal verbs
- Lists of word patterns (regular expression patterns) for sentence categorization or extracting pieces of information
  - Word patterns for extracting cause-effect information in text
  - Discourse markers for identifying rhetorical & metadiscourse functions (in genre analysis)

Resources in text mining and natural language processing (cont.)

- **Annotated corpora**
  - Social media data (product and drug reviews) annotated with sentiment categories
  - Journal articles annotated with rhetorical and metadiscourse functions, as well as information types, citation types and argument types
- **Text processing software**
  - Text segmentation (to analyse structure of documents and sentences)
  - Syntactic/semantic parsing (to identify relations between words or concepts)

WKWSCI Sentiment Lexicon

- **Manually coded by WKWSCI undergrad students over a period of 4 years**

  Total # words: ~ 30,000 words
  # positive words: ~ 3,000
  # negative words: ~ 7,000
  # neutral words: ~ 20,000
  Includes adjectives, adverbs, verbs, nouns

- **Coded on a 7-point sentiment strength scale**
  - **Very positive, Positive, Slightly Positive**
  - **Neutral**
  - **Very negative, Negative, Slightly negative**

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Evidence Based Academic Writing Assistant

version 0.1 (under development)

A free tool that assists university students in writing research papers. The tool is supported by a catalogue of linguistic, information and argument patterns derived from a corpus of articles sampled from high-impact journals. Learn more User guide Recommended: (1366 x 768)

Email:
Password:
Register Login

Rhetorical structure model ▾ Sociology ▾

Title

Lexicon-Based Sentiment Analysis: Comparative Evaluation of Six Sentiment Lexicons

Introduction

Type introduction section here...

Analyse

Literature Review

Type literature review section here...

Download Save

Introduction:
The core of the Introduction section is the research objective(s) or research question(s) statement. The text before the research objective introduces the broader research area, tradition, theory or context in which the study is situated. The research area is gradually narrowed down to the research problem/issue that the study addresses, including the specific research gap that the research objective addresses. The text will also indicate why the research issue is central, important or worth studying.

Introduction - Typical Structure — Rhetorical Functions

Move 1: Introduce the field

- Step 1: Introduce the research area/topic
- Step 2: Make a general statement about the area/topic (optional)
- Step 3: Outline the historical development of the area/topic (optional)
- Step 4: Elaborate on the research area/topic
- Step 5: Narrow down the topic (optional)

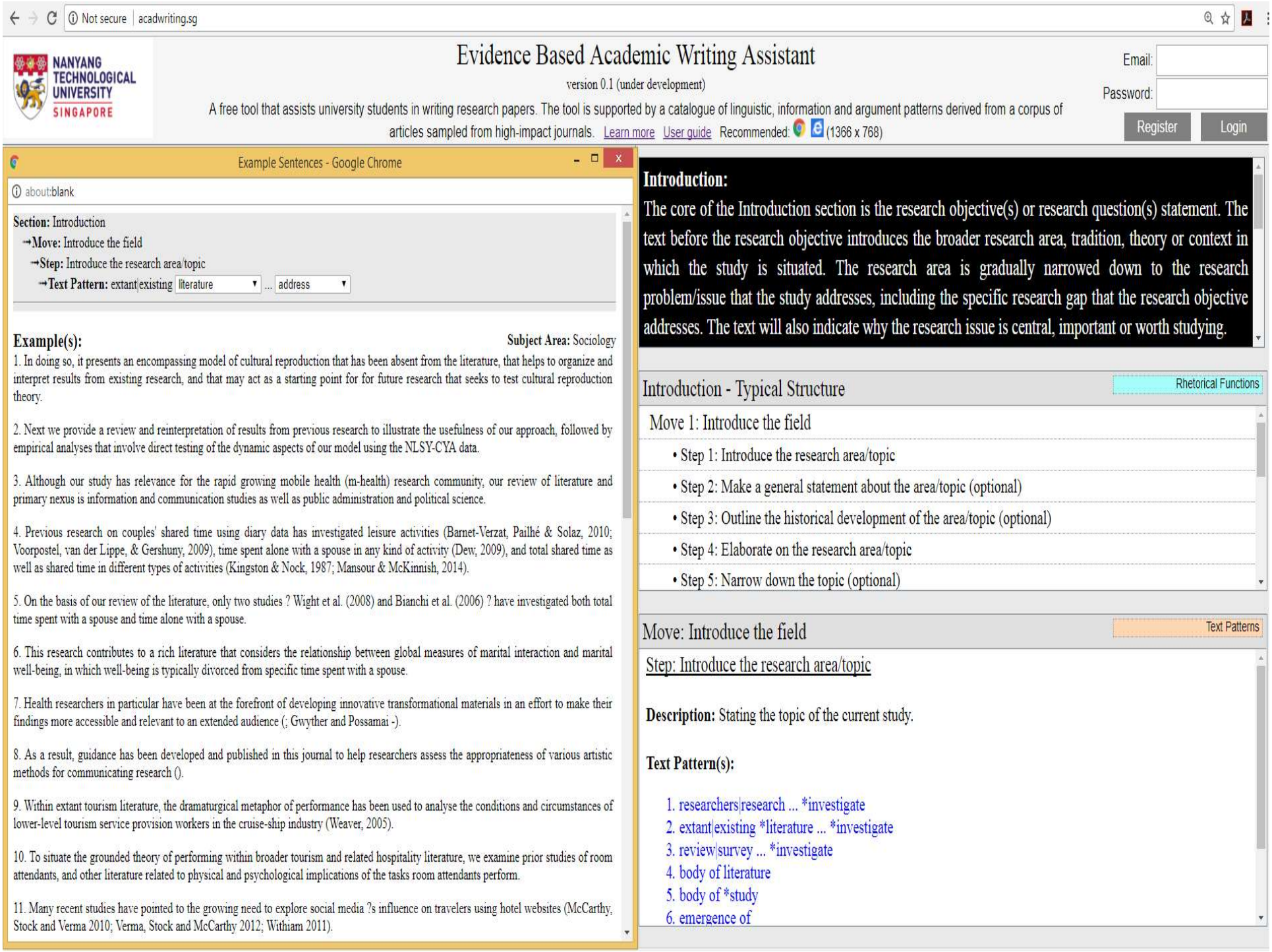
Move: Introduce the field — Text Patterns

Step: Introduce the research area/topic

Description: Stating the topic of the current study.

Text Pattern(s):

1. researchers|research ... *investigate
2. extant|existing *literature ... *investigate
3. review|survey ... *investigate
4. body of literature
5. body of *study
6. emergence of


Not secure | acadwriting.sg
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE
Evidence Based Academic Writing Assistant
version 0.1 (under development)
A free tool that assists university students in writing research papers. The tool is supported by a catalogue of linguistic, information and argument patterns derived from a corpus of articles sampled from high-impact journals. Learn more User guide Recommended: (1366 x 768)
Email:
Password:
Register
Login
Example Sentences - Google Chrome
about:blank
Section: Introduction
→Move: Introduce the field
→Step: Introduce the research area/topic
→Text Pattern: extant|existing literature ... address
Example(s):
Subject Area: Sociology
1. In doing so, it presents an encompassing model of cultural reproduction that has been absent from the literature, that helps to organize and interpret results from existing research, and that may act as a starting point for for future research that seeks to test cultural reproduction theory.
2. Next we provide a review and reinterpretation of results from previous research to illustrate the usefulness of our approach, followed by empirical analyses that involve direct testing of the dynamic aspects of our model using the NLSY-CYA data.
3. Although our study has relevance for the rapid growing mobile health (m-health) research community, our review of literature and primary nexus is information and communication studies as well as public administration and political science.
4. Previous research on couples' shared time using diary data has investigated leisure activities (Barnet-Verzat, Pailhé & Solaz, 2010; Voorpostel, van der Lippe, & Gershuny, 2009), time spent alone with a spouse in any kind of activity (Dew, 2009), and total shared time as well as shared time in different types of activities (Kingston & Nock, 1987; Mansour & McKinnish, 2014).
5. On the basis of our review of the literature, only two studies ? Wight et al. (2008) and Bianchi et al. (2006) ? have investigated both total time spent with a spouse and time alone with a spouse.
6. This research contributes to a rich literature that considers the relationship between global measures of marital interaction and marital well-being, in which well-being is typically divorced from specific time spent with a spouse.
7. Health researchers in particular have been at the forefront of developing innovative transformational materials in an effort to make their findings more accessible and relevant to an extended audience (; Gwyther and Possamai -).
8. As a result, guidance has been developed and published in this journal to help researchers assess the appropriateness of various artistic methods for communicating research ().
9. Within extant tourism literature, the dramaturgical metaphor of performance has been used to analyse the conditions and circumstances of lower-level tourism service provision workers in the cruise-ship industry (Weaver, 2005).
10. To situate the grounded theory of performing within broader tourism and related hospitality literature, we examine prior studies of room attendants, and other literature related to physical and psychological implications of the tasks room attendants perform.
11. Many recent studies have pointed to the growing need to explore social media ?s influence on travelers using hotel websites (McCarthy, Stock and Verma 2010; Verma, Stock and McCarthy 2012; Withiam 2011).
Introduction:
The core of the Introduction section is the research objective(s) or research question(s) statement. The text before the research objective introduces the broader research area, tradition, theory or context in which the study is situated. The research area is gradually narrowed down to the research problem/issue that the study addresses, including the specific research gap that the research objective addresses. The text will also indicate why the research issue is central, important or worth studying.
Introduction - Typical Structure
Rhetorical Functions
Move 1: Introduce the field
• Step 1: Introduce the research area/topic
• Step 2: Make a general statement about the area/topic (optional)
• Step 3: Outline the historical development of the area/topic (optional)
• Step 4: Elaborate on the research area/topic
• Step 5: Narrow down the topic (optional)
Move: Introduce the field
Text Patterns
Step: Introduce the research area/topic
Description: Stating the topic of the current study.
Text Pattern(s):
1. researchers|research ... *investigate
2. extant|existing *literature ... *investigate
3. review|survey ... *investigate
4. body of literature
5. body of *study
6. emergence of

Questionnaire survey data

- **Survey of the types of information shared on social networking sites by university students**

  Across 7 universities

  Nanyang Technological University, Singapore

  University of Malaya, Malaysia

  University of Pondicherry, India

  National Taiwan Normal University, Taiwan

  Kyushu University, Japan

  Mahasarakham University, Thailand

  Keimyung University, South Korea

IRB application for questionnaire survey data

What will happen to the data after research completion?

- The data will be kept permanently in a secure room by the investigators for future research use.
- In addition, aggregated data (with no subject identification information) will be uploaded to NTU digital repository for public access.

Challenges

- Funding for resource preparation
  - Development, cleaning, formatting, documentation
- IP and copyright
  - Text analysis resources are usually built on more primitive resources which may be copyrighted or have IP issues
- IRB application
  - Is painful
  - What is acceptable to the review committee?
- Resources to support reuse of the data
  - Simple charting and analysis tools for survey data
  - Software for using the resource

Thank you very much!