What are Some Research Concerns of TLP?

Alden Dima
Information Technology Laboratory
National Institute of Standards and Technology
April 16th, 2021

Outline

- Motivation
- Strawman proposal for framing TLP
- Different perspectives and potential research questions:
  1. Technical Domains
  2. Actionable/Business Intelligence
  3. Text Mining and Analysis
  4. Human–Computer Interaction
  5. Linguistics
- Closing thoughts

Technical Language Processing (TLP)

Michael P. Brundage, Thurston Sexton, Melinda Hodkiewicz, Alden Dima, and Sarah Lukens

Technical language processing: Unlocking maintenance knowledge,

Manufacturing Letters 27, 42–46, Jan. 2021. DOI: 10.1016/j.mfglet.2020.11.001

Alden Dima, Sarah Lukens, Melinda Hodkiewicz, Thurston Sexton, and Michael P. Brundage **Adapting natural language processing for technical text**,

to appear in Applied AI Letters

What Sets Apart TLP Research?

- How do we distinguish our focus with other research efforts?
- What are our boundaries?
- What kinds of questions should we be asking?
- For example, here is an example of how NLP is described:

  "The ultimate goal of natural language processing is for computers to achieve human-like comprehension of texts/languages. When this is achieved, computer systems will be able to understand, draw inferences from, summarize, translate and generate accurate and natural human text and language."

  https://www.expert.ai/blog/natural-language-processing

- Not sure that this is our goal…

NLP and TLP

- There’s no denying that NLP plays a critical role in TLP
- But TLP has a different focus than “human-like comprehension of texts/languages”

Framing TLP – A Strawman Proposal

- Focus:
  - *People working in technical roles within an organization who use dedicated computer-mediated text-based communication while participating in technical processes related to the organization's business model*
- Excludes communication that:
  - *Is not related to a technical process as part of organizational business model*
  - *Is not computer-mediated or text-based – voicemail, paperwork*
  - *Does not use a dedicated channel – emails to personal mail accounts*
- Realize that this may need to be relaxed or modified

Perspective: Technical Domains

- TLP is applicable to a variety of domains, for example:
  - **Maintenance** involves functional checks, servicing, repairing or replacing of necessary devices, equipment, machinery, building infrastructure, and supporting utilities in industrial, business, and residential installation

Wikipedia: Maintenance (technical)

Questions: Technical Domains

- How can the analysis of technical text address problems in a technical domain?
- What types of data needs to be available to address specific technical problems?
- How do business procedures need to be modified to support them?
- How do we demonstrate the value of technical text in a particular domain?

Perspective: Actionable/Business Intelligence

- Actionable intelligence is the means of obtaining concrete knowledge that permits an individual to take action based on known information

  Carroll, Christopher J. 2008. “Transforming Data and Metadata into Actionable Intelligence and Information within the Maritime Domain.” Monterey, California. Naval Postgraduate School.

- Business intelligence comprises the strategies and technologies used by enterprises for the data analysis of business information to provide historical, current, and predictive views of business operations

  Wikipedia: Business intelligence

Questions: Actionable/Business Intelligence

- How can the analysis of technical text be integrated with Observe-Orient-Decide-Act (OODA) strategies to reduce operational risks?
- How can business intelligence technologies use technical text to improve organizational decision making?

Perspective: Text Mining and Analysis

- **Text mining** is the process of deriving high-quality information from text. It involves "the discovery by computer of new, previously unknown information, by automatically extracting information from different written resources."
- **Text analysis** involves information retrieval, lexical analysis to study word frequency distributions, pattern recognition, tagging/annotation, information extraction, data mining techniques including link and association analysis, visualization, and predictive analytics. The goal is to turn text into data for analysis, via application of NLP, different types of algorithms and analytical methods.

Wikipedia: Text mining

Underlined items are things that we normally think about.

Questions: Text Mining and Analysis

- What types of information can we reliably extract from the technical text contained in information systems?
- What types of information are the most problematic?
- What affordances does technical data provide that we can leverage?
- How do we adapt NLP methods to TLP?
  - How can we best use SOTA NLP methods in TLP? When should we?
- What types of technologies and architectures should we use?

Perspective: Human–Computer Interaction

- **Human–computer interaction** (HCI) studies the design and use of computer technology, focused on the interfaces between people (users) and computers. Researchers in the field of HCI observe the ways in which humans interact with computers and design technologies that let humans interact with computers in novel ways.
- **Social computing**: there has been an explosion of social science research focusing on interactions as the unit of analysis.
- In human and computer interactions, a semantic gap usually exists between human and computer's understandings towards mutual behaviors. Ontology, as a formal representation of domain-specific knowledge, can be used to address this problem, through solving the semantic ambiguities between the two parties.

Wikipedia: Human–computer interaction

Questions: Human–Computer Interaction

- What is the best strategies for enabling human-in-the-loop TLP analyses?
  - How do we accelerate and improve the quality of annotation?
- What is the impact human/computer interaction:
  - On organizational processes?
  - On the technical text that is entered into technical information systems?
  - On the quality of TLP analyses?
- How do we design information systems that improve the organizational value of technical text?

Perspective: Linguistics

- In the **functional paradigm** a language is in the first place conceptualized as an instrument of social interaction among human beings, used with the intention of establishing communicative relationships. Within this paradigm one attempts to reveal the instrumentality of language with respect to what people do and achieve with it in social interaction.

  Wikipedia: Functional linguistics

- In sociolinguistics, a **register** is a variety of language used for a particular purpose or in a particular communicative situation

  Wikipedia: Register (sociolinguistics)

Questions: Linguistics

- How do personnel communicate to each other while performing their technical roles?
- What information gets communicated and what is omitted and why?
- How much of the omitted information is either assumed or is available elsewhere?
- Can we reliably characterize and identify technical text?
- What varieties of expression are present?
  - Can we characterize and identify them?
- How do these linguistic concerns affect the value of text information?

Closing Thoughts

- Not claiming that the proposed focus, perspectives, and research problems presented here are either complete or correct
  - But it should be clear that TLP is a multidisciplinary field that can benefit from a wide range of disciplines
- Hoping to start an ongoing discussion about the scope and goals of TLP as its own subfield that engages diverse perspectives from a variety of communities

TLP-COI on Slack

Thank you!

Alden Dima
alden.dima@nist.gov