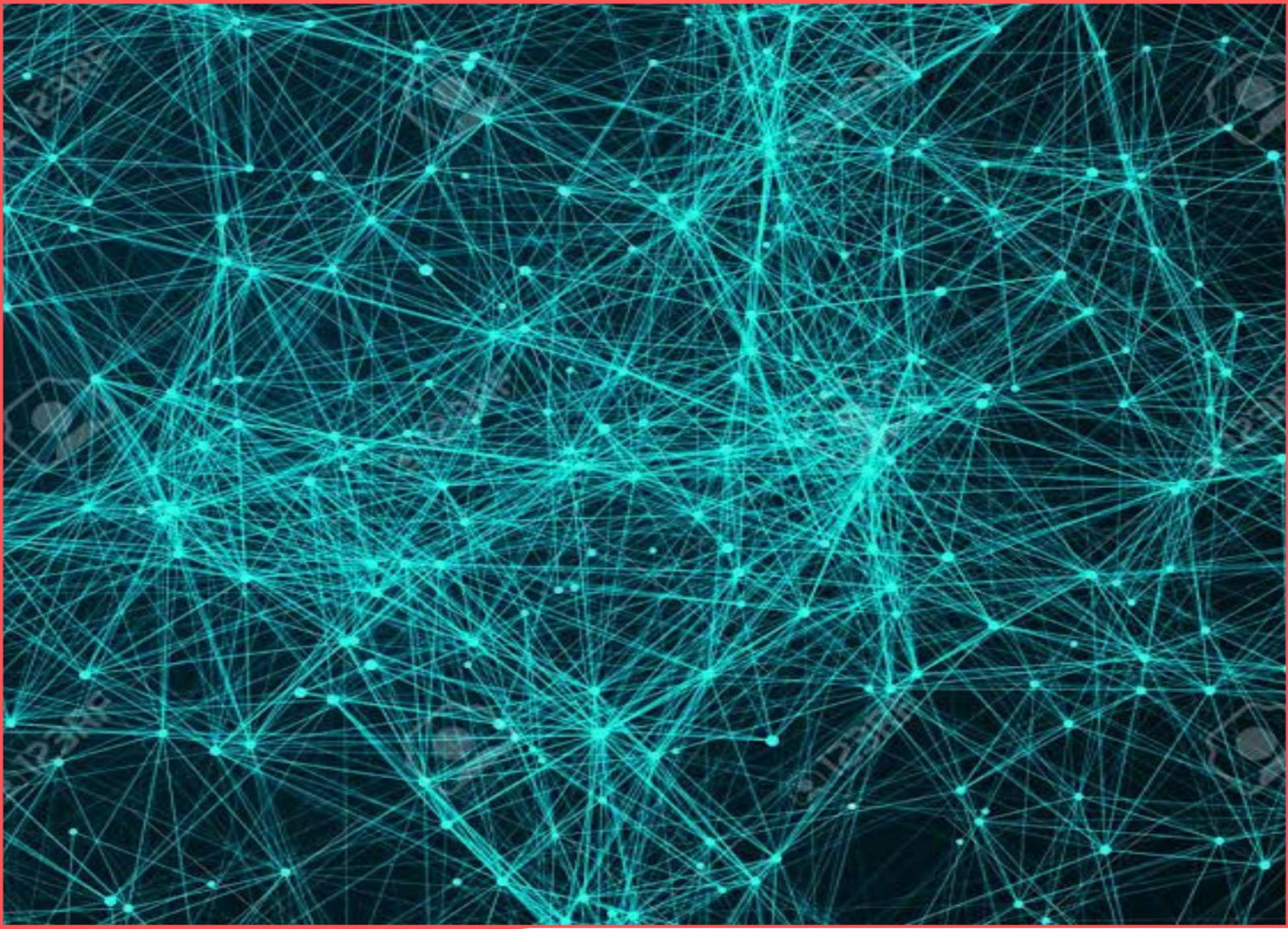
Journal of Engineering Sciences

(An open access peer reviewed Journal)



SIR PUBLISHERS
www.sirpublishers.com

ARTICLES:

INSTRUCTIONS FOR COURSE READINGS PROJECT

Prof. **John Willinsky**
Graduate School of Education
Stanford University

BACKGROUND: In Canada, university students are being charged a fixed royalty tariff to cover the royalties (due to copyright) owed to publishers and authors for their required and assigned "course readings" or "coursepacks" (made up of articles and books chapters). The tariff is based on an estimate of what they will be assigned to read for their courses. This fixed fee does not include the textbooks or lab manuals assigned in courses, which are purchased at the bookstore or online (and include the royalties in their price). In 2017, the tariff for the required course readings was $45.00 per student for one year. Is this a fair price for students to pay, in addition to the textbooks they are asked to purchase? What proportion of the course readings are journal articles and scholarly books that have already been purchased by their university library and that are the work of university faculty (as students should not be paying royalties for these)? What proportion are written and published outside of the university (as we do think it fair for students to pay royalties to use these materials)? To answer these questions, we are analyzing the course syllabuses for Canadian universities and colleges.

I. Initial notes

1. We are only interested in counting Required Readings

Whether with academic or non-academic readings, we are only counting reading materials that are "required," "assigned," "mandatory" etc. Readings are not required if they are identifiable by words like "optional," "recommended," "of interest," "helpful," etc. **If it does not say either way, then we assume it is required.** For example, a syllabus might have two sections for reading materials as here:

REQUIRED BOOKS (available at the UBC bookstore, online, and on the reserve shelf under ANTH 403H.002 at the Koerner Library)

1. The World of the Swahili: an African Mercantile Civilization. John Middleton. New Haven and London: Yale University Press.1992. **ISBN-13:** 978-0300060805

2. The Edge of Islam: Power, Personhood, and Ethno-religious Boundaries on the Kenya Coast by Janet McIntosh. Durham and London, Duke University Press. 2009. **ISBN-13:** 978-0822345091

RECOMMENDED BOOKS

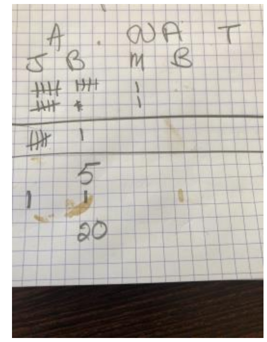
1. East Africa: An Introductory History, by Robert Maxon 3rd Revised Edition. West Virginia, The University Press. 2009. **ISBN-13:** 978-1933202464

2. Domesticating the World: African Consumerism and the Genealogies of Globalization. Prestholdt Jeremy. 2008. **ISBN-13:** 978-0520254237

3. The Swahili: The Social Landscape of a Mercantile Society by Mark Horton and John Middleton Oxford, Basil Blackwell. 2001. **ISBN-13:** 978-0631189190

4. The Swahili: Reconstructing the History and Language of an African Society, 800-1500. by Derek Nurse and Thomas Spear. University of Pennsylvania Press (January 1, 1985) **ISBN-13:** 978-0812212075

In this case, the relevant section is "Required Books" and the second section on "Recommended Books" can safely be ignored for our purposes. Most of the time there will be a "Required Reading" section and this is the section of the syllabus that is relevant for this part.



2. **Create a scratch pad with four columns**, one for academic journals, one for academic books, and for one for non-academic newspapers and magazines, one for non-academic books. This is where you can mark ticks for each syllabus and then record the total for each of the four columns in the spreadsheet.

II. What the Terms on the Data Entry Spreadsheet Mean

The **bolded** terms listed below are titles of the columns on the spreadsheet (click to have a look at your data entry spreadsheet). They are presented in the order of the columns on the spreadsheet

A. Syllabus

a. **School:** The name of the university or college.

b. **Course ID**: The course's identifying code and number (e.g., EDST465). (We provided the name of the syllabus file as a substitute in this case)

B. Select (Does not apply to the five Test Syllabuses)

a. **Inaccessible**: mark ''Yes'' if you cannot find the file you are looking for in the folder

b. **Not a syllabus**: mark anything that's not a syllabus

c. **Duplicate**: If the syllabus you open is a duplicate of one you have already analysed, that is, it is the same course from a different year or term, then simply select "Yes" and go on to the next syllabus.

d. **Other**: mark "Yes" and specify what the problem is with the file--problems including but are not limited to explicitly stating that it is from a school year before 2015

C. Academic Count

a. **Scholarly Journal Articles**: Record the number of readings on the syllabus that are scholarly journal articles.

Generally, the key indicator is that it mentions a ***volume* and/or an *issue*** number (e.g., 89/3, 52, no. 4). For example:

Licklider, Roy. 1995. The Consequences of Negotiated Settlements in Civil Wars, 1945-1993. *American Political Science Review,* 89/3: 681-690

Sambanis, Nicholas. 2000. Partition as a Solution to Ethnic War: An Empirical Critique of the Theoretical Literature. *World Politics,* 52/4: 437-483.

Here are two more, one using the title Journal, but both with volume and number: 21(3) and 99(3)

Garcia, Lorena and Lourdes Torres. 2009. "New Directions in Latina Sexualities Studies." *NWSA Journal* 21(3): pp. vii-xvi.

Kulick, Don. 1997. "The Gender of Brazilian Transgendered Prostitutes." *American Anthropologist* 99 (3), pp. 574-585.

Here are two examples that appear to be academic readings but there is no information on whether they are a journal article or book chapter. In such cases, record them as journal articles (as many chapters begin life as articles).

January 22 : **Canadian Institutionalization and the Cold War Era**

Reading: Patrice Dutil, "The Institutionalization of Foreign Affairs (1909-2013)"

Seminar Reading:

Reading: Paul Gecelovsky, "Of Legacies and Lightning Bolts: An Updated Look at the Prime Minister and Canadian Foreign Policy"

b. **Scholarly Books**: Record n**umber of books and book chapters** on the syllabus that appear to be academic or scholarly. This is a judgement call based on the following cues.

If the books or chapters are published by **university press**, or by **Routledge, Springer, Sage. Wiley, Elgar, Kluwer, Brill, Polity.** Or it has an academic-sounding title (which is more narrowly focused than a "textbook"; see textbook indicators below), although this can be tricky (e.g., *The Economic Crisis of the 1930s,* or *Aristotle*, or *Who Owns Knowledge?*). **Books have chapters, journals do not.**

Here is an example of a scholarly book and a scholarly book chapter, judging by the books' titles.

F.L. Morton and Rainer Knopff, *The Charter Revolution and the Court Party* (Peterborough: Broadview, 2000), ch. 2 (pgs. 33-58). [reserve, Isaak]

Bertha Wilson, "Decision-making in the Supreme Court of Canada," (543-6) and "Will Women Judges Really Make a Difference?" (147-152) in F.L. Morton, ed., *Law, Politics and the Judicial Process in Canada*, 3rd edition (Calgary: U of Calgary Press, 2002). [reserve, e-book via library or Isaak]

.

D. Non-Academic Count

a. **Newspapers/magazines**: Record number of readings that appear to be either a newspaper article (e.g., Globe & Mail, National Post, New York Times, etc.) or a magazine article (e.g., New Yorker, Time, Atlantic, Harper's, Macleans, Popular Mechanics, etc.).

Here are two examples of magazine articles.

Hans Koning. 1988. "Ifs: Destiny and the Archduke's Chauffeur," *Harper's* (May), pp.74-76.

Tierney, Dominic. 2011. The F-35: A Weapon That Costs More than Australia. *The Atlantic,* March 15.

b. **Non-Academic Books**: Record number of books and book chapters that appear to be non-academic, whether novels (Handmaid's Tale) or nonfiction (The Tipping Point) or poems or autobiographies.

Here are four examples of book chapters (one is a video, which we do not count). The first three are from nonacademic books, judging by the book title. A further clue for the first three is that the **readings are from a humanities course** (ENG 6181HF Permaculture and Literature), where **non-academic books are most commonly found.** The last reading (Odum) is from an academic book, as Routledge is identified above as an academic publisher.

Aldo Leopold, "Thinking Like a Mountain" from *Sand County Almanac*
Wendell Berry, "Solving for Pattern" from *The Art of the Commonplace,* 267-275
Robin Wall Kimmerer, "Asters and Goldenrod," from *Braiding Sweetgrass,* pp. 39-47
Watch: https://biomimicry.org/what-is-biomimicry/
Eugene Odum, *Ecological Vignettes* (Routledge, 2012), Chapter 26: "Ecology: The Common Sense Approach." https://books-scholarsportal-info.myaccess.library.utoronto.ca/uri/ebooks/ebooks4/taylorandfrancis4/2018-06-02/4/9781134414703

Here's an example that could go either way: A book that is written by an academic but rate it non-academic because it is published by **Penguin, a non-academic publisher, with a non-academic title.**

Pinker, Steven. 2011. *The Better Angels of Our Nature: Why Violence Has Declined.* Penguin Books: New York.

E. Textbooks

Textbook: An academic book where the whole book (and not just selected readings or chapters) is assigned and students are expected to purchase the book. It is typically a "textbook" **if the syllabus states it is the course "textbook" or a "required text"**; if it is the 3rd or higher edition of a book (it includes "3rd ed.," or "4th edition," or "Fifth edition" in the bibliographic citation of the textbook); if it has a textbook-like title, which can be tricky to tell (e.g., *Principles of Economics,* or *Calculus,* or *Organic Chemistry,* or *Handbook of International Law,* etc.) and/or it is published by the major textbook publishers: Prentice Hall, Pearson, McGraw Hill, or Cengage. Another thing to note here is that a **Laboratory Manual** may be required reading but you should ignore these in your analysis.

In this example we see the *title* of the textbook (Organic Chemistry) and the **edition** (12th).

Textbook: ***Organic Chemistry*** by T. W. Graham Solomons, Craig B. Fryhle, *and Scott A. Synder12th edition*
Laboratory Manual: Organic Chemistry 261 Laboratory Manual edition 2016
Optional Materials: Study Guide and Solution manual and molecular model kit.

In the following example, it is a little less obvious that the required readings that are listed are textbooks but from context (the preceding paragraph) it is still pretty clear:

Required Texts

This course requires the following books. Please avoid purchasing the incorrect edition. The books are available at the Carleton University Bookstore or can be purchased online at major retailers. You are welcome to purchase e-editions or second hand copies if available:

- Steven Lamy et al. Introduction to Global Politics. Fourth Edition. Oxford UP, 2016
- Jan Klabbers. International Law. Cambridge UP, 2013

In this example, the only required reading is a single textbook and everything else is listed as recommended (which we are not interested in).

Textbooks:
Required: Insel, P., Roth, W., Irwin, J., & Burke, S. (2012) *Core Concepts in Health: Canadian edition.* McGraw-Hill Ryerson

Recommended: Hacker, D. & Sommers, D. (2011) ***A Canadian Writer's Reference, 5th Ed***. Scarborough, Ont., Thomson Canada, or a similar writers reference.

Here is another example with 2 required textbooks, as well as recommended books which we do not want to count.

Textbooks:

1- Jacob, D. J., Introduction to atmospheric chemistry, Princeton University Press, 1999. [Available online and for download at: http://acmg.seas.harvard.edu/people/faculty/djj/book/index.html].
2- Schnelle, K. B.; Brown, C. A, Air pollution control technology handbook, CRC Press, 2002. [Available electronically through the library, only two simultaneous users]

Note that there will be additional material in the lectures. Your main source shall be your class notes.

Recommended books:

1- Seinfeld, J. H.; Pandis, S. N., Atmospheric chemistry and physics: from air pollution to climate change, Wiley-Interscience, 2006 (2^{nd} ed).
2- De Nevers, Noel, Air pollution control engineering by, McGraw-Hill, 2000 (2^{nd} ed).
3- Wark, K.; Warner, C. F.; Davis, W. T., Air Pollution: Its origin and control, Addison Wesley, 1998 (3^{rd} ed).

III. Notes on What NOT to Count and How to Count
updated: 6/27/2017

***What to ignore in your analyses:**

lab manuals
course packs/course guides
mentions of readings in course websites/supplemental online readings (but you don't know what the

name of the reading is)
basically all NGOs and IGOs and the UN and gov't docs-- example. policy documents (think govt speeches), reports
documentaries
readings that are identified as "resources" or "bibliography" or even "references" - we do not consider them required readings - but if they do not fall under those headings and are not explicitly stated as "required" or "optional", the default is to count them

***Do I count readings mentioned with full names and are available on "course reserves"?** - YES

***What if a syllabus mentions a link/course website for reading?**
If a link to access the reading is provided, but it's broken, should we count it as one reading?--The fact it's being provided via a link (or is on the course website) does not determine whether we count it or not.

More important is whether we know what the reading is. if we don't know what the reading is (and they say we will find out on the course website), then we ignore that one.

But if we know the name of the reading (and they say it's accessible via website or give us a link to access), we still count the reading based on its type, regardless of whether the link is broken or the website is inaccessible.

***Do I count a chain-listing of different pages as 1 reading or many?**
-If it is an assigned reading with different sections and page numbers identified then it is only one reading.
Example: Penner and Ferdinand (2009) Overcoming Katrina: African American voices from the Crescent City and beyond, London: Palgrave. Read "Introduction" (pp. xvii-xxv) and at least four narratives, one from each of the book's four main sections: (i) Retirees; (ii) At the height of their careers; (iii) Thirty somethings; (iv) Coming of age. (e-book)
-If multiple chapters or sections are assigned at different points in the course (and it is not a designated "textbook"), then each assigned reading counts as an academic book (chapter).
-But if a book is designated as "textbook" or "required text" and pops up at different points in the course with different pages, then only count them once and as textbooks.

Example: "Required Texts: (Available at Octopus Books, 116 Third Ave. Ottawa, ON K1S 2K1): Dangarembga, Tsitsi. Nervous Conditions"

Feb 28	Dangarembga, *Nervous Conditions* Chapters 1-5
Mar 6	Dangarembga, *Nervous Conditions* Chapters 6-10

INTRODUCTION TO COLLABORATIONS IN RESEARCH INFRASTRUCTURES

Author: **Jennifer Edmond**

KEYWORD(S):
Open Science; Research Infrastructures; managing and sharing research data data management planning; RDM; DMP; digital humanities; collaboration; interdisciplinary; interdisciplinarity

Text for index page:

PARTHENOS

Archived snapshot of the Introduction to Collaborations in Research Infrastructures module, which is part of the PARTHENOS Training suite [1], which was developed as part of Work Package 7 in the PARTHENOS project [2].

By the end of this module, learners should be able to:

- Understand what is meant by collaboration in humanities research
- Be aware of how this model impacts upon the development of digital humanities, and digital humanities research infrastructures

Background:

The PARTHENOS project [3] recognised that over the past ten years, researchers, institutional leaders and policymakers have begun to speak more and more about infrastructure. As more voices join the conversation, however, it can sometimes become more difficult, rather than less, to understand what exactly research infrastructure is and does. In particular in the humanities, and the digital humanities, the term is used to cover a lot of different projects, resources and approaches.

To address this gap, the PARTHENOS cluster of humanities research infrastructure projects devised a series of training modules and resources for researchers, educators, managers, and policy makers who want to learn more about research infrastructures and the issues and methods around them.

The modules, which have been released on a rolling basis from late 2016, cover a wide range of awareness levels, requirements and topic areas within the landscape of research infrastructure.

This deposit is never intended to replace the online version of the training material on the PARTHENOS website, and is intended as an archive of content.

[1] https://training.parthenos-project.eu/

[2] WP7 – Skills, Professional Development and Advancement: http://www.parthenos-project.eu/resources/projects-deliverables#1523355756261-be477222-2866

[3] http://www.parthenos-project.eu/

[This is an archived snapshot of an online course. The online course may be updated over time, and though new versions will be created to reflect major changes, the archived version may not match exactly the content of the online version]

TEXT ANALYSIS USING DEEP NEURAL NETWORKS IN DIGITAL HUMANITIES AND INFORMATION SCIENCE

Author: **Kristina Edmond**

ABSTRACT:
Combining computational tehischnologies and humanities is an ongoing effort aimed at making resources such as texts, images, audio, video, and other artifacts digitally available, searchable, and analyzable. In recent years, deep neural networks (DNN) dominate the field of automatic text analysis and natural language processing (NLP), in some cases presenting a super-human performance. DNNs are the state-of-the-art machine learning algorithms solving many NLP tasks that are relevant for Digital Humanities (DH) research, such as spell checking, language detection, entity extraction, author detection, question answering, and other tasks. These supervised algorithms learn patterns from a large number of "right" and "wrong" examples and apply them to new examples. However, using DNNs for analyzing the text resources in DH research presents two main challenges: (un)availability of training data and a need for domain adaptation. This paper explores these challenges by analyzing multiple use-cases of DH studies in recent literature and their possible solutions and lays out a practical decision model for DH experts for when and how to choose the appropriate deep learning approaches for their research. Moreover, in this paper, we aim to raise awareness of the benefits of utilizing deep learning models in the DH community.

INTRODUCTION

The research space of digital humanities (DH) applies various methods of computational data analysis to conduct multi-disciplinary research in archaeology (Eiteljorg, 2004; Forte, 2015), history (Thomas, 2004; Zaagsma, 2013), lexicography (Wooldridge, 2004), linguistics (Hajic, 2004), literary studies (Rommel, 2004), performing arts (Saltz, 2004), philosophy (Ess, 2004), music (Burgoyne, Fujinaga, & Downie 2015; Wang, Luo, Wang, & Xing, 2016), religion (Hutchings, 2015) and other fields. The scope of DH continues to expand with the development of new information technologies, and its boundaries remain amorphous (McCarty, 2013). Therefore, DH's definition is unclear and may have different interpretations (Ramsay, 2016; Poole, 2017). Library and Information Science (LIS) and DH research have a similar and overlapping scope and interfaces (Posner, 2013; Koltay 2016), to the extent that some propose to integrate and combine both research fields (Sula, 2013; Robinson, Priego, & Bawden, 2015). DH and LIS academic units are often located together (Sula, 2013), and share a significant volume of common topics, such as metadata, linked data and ontologies, information retrieval, collection classification, management, archiving and curation, bibliographic catalogue research, digitization of printed or physical artifacts, preservation of cultural heritage, data mining and visualization, and bibliometrics (Svensson, 2010; Russell 2011; Gold, 2012; Warwick 2012; Sula, 2012; Beaudoin, & Buchanan, 2012; Sula 2013; Drucker, Kim, Salehian, Bushong, 2014; Koltay, 2016; Gold, & Klein 2016). However, regardless of the definition or research scope, many (if not most) of the research in DH/LIS focuses on textual resources, recorded information, and documents (Robinson et al., 2015; Poole, 2017). Therefore, this paper argues that a deep understanding of text analysis methods is a fundamental skill that future (and present) DH/LIS experts must acquire.

Supervised deep neural networks (deep learning) are a subset of machine learning algorithms considered to be the state-of-the-art approach for many NLP tasks, such as entity recognition (Li, Sun, Han, & Li, 2020), machine translation (Yang, Wang, & Chu, 2020), part-of-speech tagging and other tasks (Collobert & Weston, 2008) from which many DH/LIS text analysis research projects can benefit. Therefore, this paper aims to raise the awareness of DH and LIS researchers of state-of-the-art text analysis (NLP using deep neural networks) approaches and techniques. This is not the first attempt to make NLP technologies accessible or highlight the benefits of NLP to the DH/LIS research community (Biemann, Crane, Fellbaum, & Mehler, 2014; Kuhn, 2019; Hinrichs, Hinrichs, Kübler, & Trippel, 2019; McGillivray, Poibeau, & Ruiz Fabo, 2020). However, this paper argues that in addition to bridging between the NLP community and the DH/LIS research community, the DH/LIS research community should cultivate experts with a deep understanding of the technological space, experts that are capable of customizing and developing the technology themselves. Use of "off the shelf" tools and algorithms is no longer sustainable (Kuhn, 2019); the future DH expert must be comfortable using and adapting state-of-the-art NLP methodologies and technologies to the DH-specific tasks. To the best of our knowledge, this is the first attempt to highlight the challenges and analyze the potential solutions of the common usage of deep neural networks for text analysis in the DH/LIS space.

DNN models are often developed by computer scientists and trained, tested, and optimized for generic, open-domain tasks or by commercial enterprises for modern texts (Krapivin, Autaeu, & Marchese, 2009; Rajpurkar, Zhang, Lopyrev, & Liang, 2016). However, applying these DNN models for DH/LIS tasks and textual resources is not straightforward and requires further investigation. This paper presents the practical challenges that DH/LIS experts may encounter when applying DNN models in their research by examining multiple use cases presented in current literature, alongside an overview of the possible solutions, including deep learning technology. Although there might be other methodological challenges (Kuhn, 2019), this paper focuses on the two main practical challenges faced when applying deep learning for almost every DH research:

(1) Training data (un)availability - DH text resources are often domain-specific and niche, and contain a relatively small number of training examples; thus, there is not enough data for the DNN learning process to converge. Even when there is a large DH text corpus, there are no balanced ground truth labeled datasets (i.e., datasets with the distribution of "right" and "wrong" examples representative of the corpus) from which the DNN can learn (McGillivray et al., 2020), and changes or adaptations in the network architecture are required in order to achieve high accuracy for such datasets (Hellrich & Hahn, 2016).

(2) Domain adaptation - in many tasks considered "common" in NLP, the DH interpretation of the task is different from the standard interpretation. Moreover, DH text resources may need to be preprocessed before serving as input to DNNs, due to "noisy" data (biased, contains errors or missing labels or data (Hall, 2020; Prebor et al., 2018)) or non-standard data structure, such as mixed data

formats (combining unstructured text, semi-structured and structured data in the same resource). In many cases, these resources are unsuitable for serving as an input into DNN models, or if they are used as-is, the models do not achieve maximum accuracy.

These challenges have unique implications on the utilization of DNNs with DH/LIS resources and tasks and, in various cases, may require different solutions. As a result of this study, a decision model for choosing the appropriate machine-learning approach for DH/LIS research is presented as a practical guideline for experts, with topics that digital humanitists should master being outlined.

Digital Humanities and Automatic Text Analysis

Natural Language Processing (NLP) is a research area that explores how computational techniques (algorithms) can be used to understand and transform natural language text into structured data and knowledge (Young, Hazarika, Poria, & Cambria, 2018; Chowdhary, 2020). Until a few years ago, the state-of-the-art techniques that addressed supervised natural language processing challenges were based on a mix of machine learning algorithms. NLP tasks such as text classifications, entity recognition, machine translation, and part-of-speech tagging were solved using various classic supervised machine learning algorithms, such as Support Vector Machine (SVM), Hidden Markov Model (HMM), decision trees, k-nearest neighbors (KNN), and Naive Bayes (Zhou & Su, 2002; Liu, Lv, Liu, & Shi, 2010; Vijayan, Bindu, & Parameswaran, 2017). Basically, these algorithms apply a manually selected set of characteristic features to a given task and corpus, and a labeled dataset with "right" and "wrong" examples for training the optimal classifier. Given a new example of the same type, this classifier will be able to automatically predict whether or not this example belongs to the predefined category (e.g., whether a given sentence has a positive sentiment or not).

However, in many cases, it is not easy to decide what features should be used. For example, if a researcher wishes to learn to classify a text's author from the Middle Ages, she will need to use the features that represent the unique writing styles that distinguish the authors. Unfortunately, it is not easy to describe these features in terms of textual elements. Deep learning solves this central problem by automatically learning representations of features based on examples instead of using explicit predefined features (Deng & Liu, 2018). Deep learning (DL) is a sub-field of machine learning that draws its roots from the Neurocognition field (Bengio, Goodfellow, & Courville, 2017). The DL approach uses deep neural networks (DNN) models for solving a variety of Artificial Intelligence tasks. The technical details of various DNN models and techniques appear in Appendix I.

DH researchers use NLP algorithms for DH-specific tasks in various domains. For example, Niculae, Zampieri, Dinu, and Ciobanu (2014) used NLP techniques to automatically date a text corpus. They developed a classifier for ranking temporal texts and dating of texts using a machine learning approach based on logistic regression on three historical corpora: the corpus of Late Modern English texts (de Smet, 2005), a Portuguese historical corpus (Zampieri & Becker, 2013) and a Romanian historical corpus (Ciobanu, Dinu, Dinu, Niculae, & Sulea, 2013). To construct social networks among literary characters and historical figures, Elson, Dames, and McKeown (2010) applied "off-the-shelf" machine learning tools for natural language processing and text-based rules on 60 nineteenth-century British novels. Zhitomirsky-Geffet and Prebor (2019) used lexical patterns for Jewish sages disambiguation in the Mishna, and then applied several machine learning methods based on Habernal and Gurevych's (2017) approach for the co-occurrence of sages and pattern-based rules for specific inter-relationship identification in order to formulate a Jewish sages social interactions network. In paleography, the study of historical writing systems and the deciphering and dating of historical manuscripts, Cilia, De Stefano, Fontanella, Marrocco, Molinara, and Freca (2020) utilized MS-COCO (Lin, Maire, Belongie, Hays, Perona, Ramanan, & Zitnick, 2014), a generic corpus of images, and a domain-specific corpus to train DNN models and design a pipeline for medieval writer identification. To predict migration and location of manuscripts, Prebor, Zhitomirsky-Geffet and Miller (2020a, 2020b) devised lexical patterns for disambiguation of named entities (dates and places) in the corpus of the Department of Manuscripts and the Institute of Microfilmed Hebrew Manuscripts in the National Library of Israel. Next, the authors trained a CART machine learning classifier (Classification and regression tree based on Decision Tree learning) (Rokach and Maimon, 2015) to predict the places of manuscripts that were often absent in the corpus. For ancient languages analysis, a study (Dereza, 2018) compared accuracy for lemmatization for early Irish data using a rule-based approach and DNN models, and proved the advantages of using DNN on such a historical language - even with limited data. For historical network analysis, Finegold, Otis, Shalizi, Shore, Wang, and Warren (2016) used named entity recognition tools (Finkel, Grenager, & Manning, 2005; Alias-i, 2008) with manual rules on the Oxford Dictionary of National Biography and then applied a regression method, namely Poisson Graphical Lasso (Yang, Ravikumar, Allen & Liu, 2013) to find correlations between entities (nodes). Nevertheless, as demonstrated by the examples above, although there is a "computational turn" (Berry, 2011) in the DH research and methodologies, state-of-the-art computational NLP algorithms, like deep neural networks, are still rarely used within the core research area of DH (Kuhn, 2019).

To estimate the potential of deep learning use in DH, a comparison has been performed to one of the fields that is similar to DH - Bioinformatics. These fields are comparable since both are characterized by their inter-disciplinarity and because Bioinformatics thrives on application of computational analysis for exploring and investigating information repositories in a chosen knowledge domain (Ewens & Grant, 2006). A list of leading journals was compiled in each field and searched for articles with "deep neural network" and "machine learning" keywords. For DH, twelve journals were selected, based on Spinaci, Gianmarco, Colavizza, Giovanni, & Peroni (2019), all in English and ranked as 1 (exclusively DH). For Bioinformatics, twelve journals were selected based on Google Scholar's top publication list[1]. The two lists of the journals appear in Appendix III.

The comparison was conducted on the articles published in the above journals over the past three years and measured the following: 1) the percentage of articles with each of the two keywords in the selected journals in each field, to ascertain the usage of machine learning (ML) in general vs. deep learning (DL) in particular, in each field; and 2) the percentage of articles mentioning deep learning out of the machine learning articles in each field. As can be observed from Figure 1, in the DH field,

[1] https://scholar.google.com/citations?view_op=top_venues&hl=en&vq=bio_bioinformatics

only 21% of the articles discussing "machine learning" also discussed "deep learning"; while in Bioinformatics, 52% of the articles discussing "machine learning" also discussed "deep learning". Moreover, in the DH field, only 3.8% of the articles mentioned "deep learning", while in Bioinformatics, 19.5% of the articles mentioned "deep learning" – five times higher. In addition, in the DH field, 18% of the articles discussed "machine learning", while in Bioinformatics, 37% of the articles discussed "machine learning" – only two times higher. These results indicate that the DH field "lags behind" when it comes to using machine learning and especially deep learning state-of-the-art models.

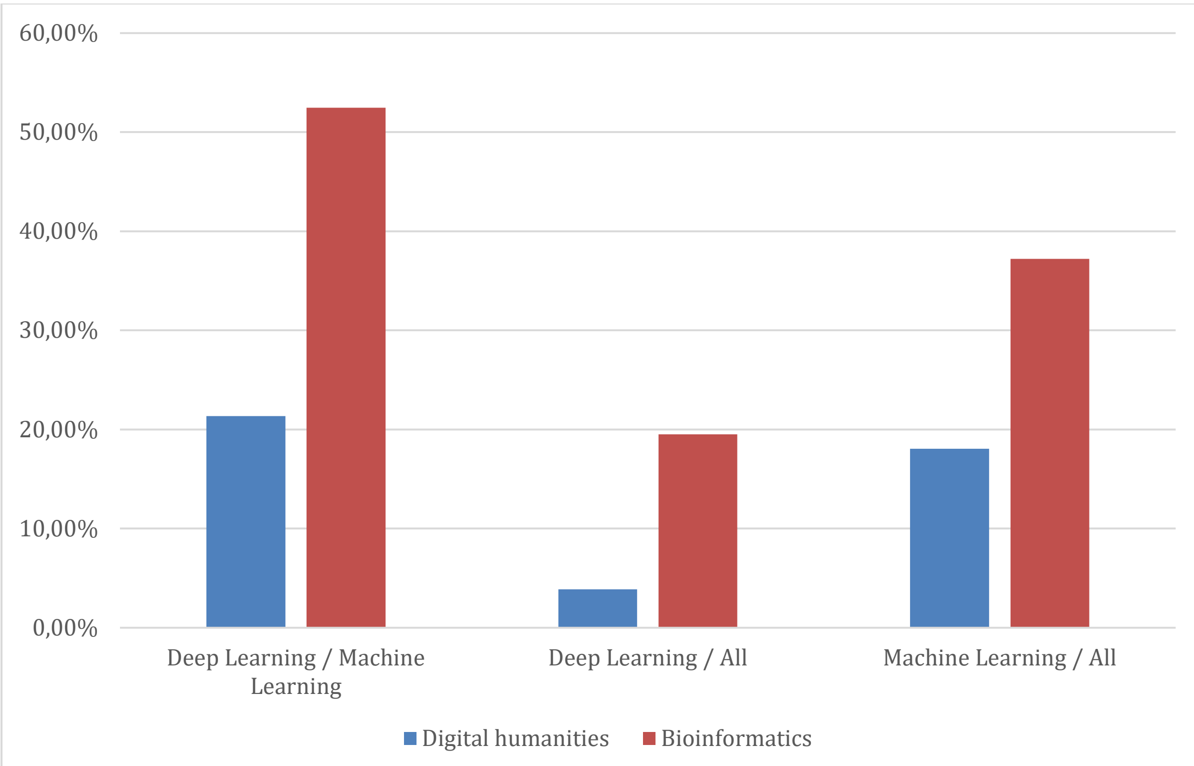


Figure 1: Deep neural networks and machine learning articles in DH/LIS vs. Bioinformatics.

The next section provides an in-depth analysis of challenges and potential solutions for using DNN in DH/LIS, supported by multiple use-case studies from the recent DH literature. The analysis is divided into two main sections dealing with two primary challenges in applying deep learning to DH research: training data (un)availability and domain adaptation.

Challenges when Using Deep Learning for Digital Humanities Research

Training Data (Un)availability

Computer scientists often work on generic supervised text analysis tasks with open-domain or modern datasets. Kaggle[2], the machine learning community, hosts many of these datasets. For example, the IMDb dataset contains a short description of a movie, and its review score allows to research sentiment analysis (Maas, Daly, Pham, Huang, Ng, & Potts, 2011); question answering system can be developed using Stanford Question Answering Dataset (Rajpurkar, Jia, & Liang, 2018);

[2] https://www.kaggle.com/datasets

and SPAM filtering can be developed using a dedicated dataset (Almeida, Hidalgo, & Silva, 2013). Unfortunately, the DH community has not (as yet) produced large annotated open datasets for researches (although there are few in niche areas like (Rubinstein, 2019; Chen & Chang, 2019)). The lack of annotated data is a challenge for both classical machine learning and deep learning supervised algorithms (Elmalech & Dishi 2021). However, supervised deep learning algorithms require significantly more data than machine learning algorithms, making this challenge a critical practical challenge for DH researchers. This is one reason that even when DH/LIS researchers use deep learning, they often use unsupervised algorithms that do not require training data and are limited to specific tasks (Moreno-Ortiz, 2017). This section investigates some of the methods that DH researchers can apply to overcome this challenge.

Training Dataset Generation by Humans

Humans are the best alternative for dataset generations due to their domain knowledge and high accuracy. Therefore, the first consideration when generating a dataset is to consider if humans can be used for the job. However, humans are not as scaleable as computer software. It is possible to manually generate a dataset by humans when the needed labeling is relatively small or as a baseline for synthetic dataset generation. There are two types of manual dataset generation: crowd-based dataset generation and domain expert-based dataset generation. Crowdsourcing dataset generation is a relatively cheaper and effective method, but it can only be used when the labeling is "common knowledge". In some cases, for example, in the study aiming to generate a dataset of relationships extraction between characters in literary novels (Chaturvedi et al., 2016), the researchers must use expert annotators that can read and understand a novel; or even annotate themselves when working with historical languages known only to a few, as in Schulz & Ketschik (2019).

Crowdsourcing is based on large groups of non-expert, low-paid workers or volunteers performing various well-defined tasks. Existing studies tested optimization strategies for different tasks, such as extracting keyphrases (Yang, Bansal, Dakka, Ipeirotis, Koudas, & Papadias, 2009), natural language and image annotation (Snow, O'Connor, Jurafsky, & Ng, 2008; Sorokin & Forsyth, 2008), and document summarization (Aker, El-Haj, Albakour, & Kruschwitz, 2012). Crowdsourcing requires quality control to ensure that crowd workers are performing their tasks at a satisfactory level (Elmalech Grosz 2017). One of the effective generic (task-agnostic) quality control techniques is filtering out tasks with a low inter-worker agreement (Bernstein, Little, Miller, Hartmann, Ackerman, Karger, Crowell, & Panovich, 2010; Downs, Holbrook, Sheng, & Cranor, 2010; Kittur, Smus, Khamkar, & Kraut, 2011). Another popular approach is breaking tasks into sub-tasks (Bernstein et al., 2010; Kittur et al., 2011).

Employing crowd workers for dataset generation has been carried out in various domains, including DH projects (e.g., Elson, Dames, & McKeown, 2010). Thus, in this use-case study, Elson et al. (2010) utilized crowdsourcing to build a dataset of quoted speech attributions in historical books in order to generate a social network among literary characters. Elson et al. (2010) did not use DNN, but rather classic machine learning methods (Davis, Elson, & Klavans, 2003), but the dataset generating process is the same for classic ML and DL.

Another example of such a use-case is fixing Optical Character Recognition (OCR) errors in historical texts. In the DH/LIS space, there is great interest in investigating historical archives. Therefore, over the past few decades, archives of paper-based historical documents have undergone digitization using OCR technology. OCR algorithms convert scanned images of printed textual content into machine-readable text. The quality of the OCRed text is a critical component for the preservation of historical and cultural heritage. Unsatisfactory OCR quality means that the text will not be searchable, analyzable, or analysis may result in wrong conclusions. Unfortunately, while generic OCR techniques and tools achieve good results on modern texts, they are not accurate enough when applied to historical texts. Post-correction of digitized small scale or niche language historical archive is a challenge that can be solved using DNNs with high accuracy (Chiron, Doucet, Coustaty, & Moreux, 2017; Rigaud, Doucet, Coustaty, & Moreux, 2019) if an appropriate dataset is attainable. Therefore, the first thing that should be researched is an effective methodology for crowdsourcing this specific task (Suissa, Elmalech, & Zhitomirsky-Geffet, 2019). The details of the crowdsourcing research are outside the scope of this paper. What is essential from the DH/LIS research point of view is that the findings of Suissa et al. (2019) proved to be an effective dataset generation approach. Using the developed strategies, DH researchers can optimize the process to achieve better results matching their objectives and priorities. The corrected corpus of OCRed texts created by the optimized crowdsourcing procedure can serve as a training dataset for DNN algorithms.

However, although the crowdsourcing method yields satisfactory results, it is suitable mainly for widely spread languages like English or Spanish. Other national languages do not have enough crowd workers-speakers to utilize such an approach effectively. Moreover, manually generating a dataset for training a DNN model in order to post-correct OCR errors is expensive and inefficient, even when the task is crowdsourced. Therefore, in practice, this human-only dataset generation should be shifted to a human-in-the-loop solution.

Training Dataset Generation using Algorithms

The next range of solutions takes a two-phase approach. In the first phase, humans are used to create a small set of examples; this set of examples is used in the second phase by a different set of algorithms to generate a synthetic dataset with numerous training examples (Pantel, & Pennacchiotti, 2006; Bunescu, & Mooney, 2007). One way is to find recurring patterns in a small number of manually corrected examples, and use them to generate more correct examples. Thus, the use-case study that adopted this approach for automatic training dataset generation in the OCR post-correction domain, Suissa, Elmalech, & Zhitomirsky-Geffet (2020) used crowd workers to fix a relatively small set of OCRed documents. Then, the Needleman–Wunsch alignment algorithm (Needleman, & Wunsch, 1970) was used to find common confusions between characters committed by the crowd workers. Using this confusion list, a large dataset of "wrong" and "right" sentences was generated and used by a DNN to correct historical OCRed text.

Another way to generate a dataset from a small set of manual examples is called "distant supervision" (Mintz, Bills, Snow, & Jurafsky, 2009). In this approach, a classifier is trained on a small set of examples and is applied to a large corpus. The classifier will classify the data with a relatively low accuracy but sufficiently high accuracy for the DNN to learn other features from this weak classification. Blanke, Bryant, & Hedges (2020) used this method to perform sentiment analysis on Holocaust testimonials data (Thompson, 2017). In the first phase, they did not use crowd workers for the initial dataset generation but rather applied a dictionary-based approach to find negative and positive sentiment sentences based on the TF-IDF measure (Singhal, 2001). Using these sentences, they trained a classifier to distinguish between positive and negative examples. In the second phase, they used the classifier to produce a large training corpus of positive and negative memories of Holocaust survivors for DNN text analysis. Using this method eliminates the need for humans; however, it is suitable only for specific tasks.

A different approach to solving the training dataset's unavailability is the transfer learning (Torrey, & Shavlik, 2010) method. In transfer learning, a generic dataset is used; the dataset should be suitable for the task needed to be solved, but with open-domain / other domain data. The model is then trained again using a small set of domain-specific examples (generated by humans or artificially). This approach is based on the intuition that humans transfer their knowledge between tasks based on previous experiences. Cilia et al. (2020) utilized transfer learning to identify medieval writers from scanned images. Instead of generating a large dataset, they used a model that was already trained on an open generic dataset MS-COCO (Lin et al., 2014) and trained it again using a small set of domain-specific examples from the Avila Bible (images of a giant Latin copy of the Bible). Banar, Lasaracina, Daelemans, & Kestemont (2020) applied transfer learning to train neural machine translation between French and Dutch on digital heritage collections. They trained several DNNs on Eubookshop (Skadiņš, Tiedemann, Rozis, & Deksne, 2014), a French-Dutch aligned corpus. Then, instead of training the DNN models directly on the target domain data, they first trained the models on "intermediate" data from Wikipedia (articles close to the target domain). Only then did they train the models for the third time on the target domain data - the Royal Museums of Fine Arts of Belgium dataset. Using this "intermediate fine-tuning" approach, Banar et al. (2020) achieved high accuracy for French-Dutch translation in the domain of Fine Arts. This method can also solve another challenge for the DH/LIS researcher when using DNN models – the domain adaptation challenge.

Recent studies (Radford, Wu, Child, Luan, Amodei, & Sutskever, 2019; Brown, Mann, Ryder, Subbiah, Kaplan, Dhariwal, & Amodei, 2020) show that in some cases, instead of fine-tuning a pre-trained model, a large-scale pre-trained model, such as GPT3 (Radford et al., 2019), trained on ~500 billion (modern) words, can achieve good results with a limited (or without) domain-specific dataset. Although these methods (named Few-shot and Zero-shot learning) do not reach the same performance as the fine-tuning method, they are preferable for low resource domains when dataset generation is impossible. However, most of the models that are pre-trained on a large-scale modern English dataset and suitable for Few-shot and Zero-shot learning may not reach the same accuracy for DH historical corpora, especially in (other than English) national languages, due to a bias towards modern language.

Domain Adaptation

Even with a large dataset ready for DNN training, there are other challenges a DH/LIS expert may encounter when attempting to solve a text analysis task on DH/LIS data with DNNs. As mentioned in the previous section, data is a critical part of DNN's high accuracy. However, specific task/domain adaptation is just as vital, and without adapting the model or the architecture to the specific task and domain, the DNN may perform poorly.

A DNN model is a set of chained mathematical formulas with weights assigned to each node (neuron) expressing a solution to a specific task. Although there are regulation techniques to generalize the DNN model, in many cases training the model with different data will significantly impact the weights. In other words, using the same mathematical formulas, the learning process interprets the same task differently. In this context, transfer learning described in the previous section can also serve as a domain adaptation method, since the DNN model's interpretation of the task is adjusted to the domain-specific data. Moreover, DH/LIS text analysis tasks are not just different in terms of interpretation but also often require domain-specific preprocessing and analysis pipeline. Therefore, in order to improve the accuracy of DNN models for text analysis tasks, DH/LIS experts should be familiar with methods and techniques for customizing DNN models, preprocessing DH/LIS data, and adapting the analysis pipeline.

DNN Optimization for DH-specific Tasks

A DNN model has a high number of architecture components and hyper-parameters that influence the model training efficacy and accuracy. Selecting the domain-specific suitable components and hyper-parameter values may considerably improve the performance of the DNN (Bengio, 2012). Here are a few of the most common architectures and hyper-parameters that an expert should consider (see Appendix I for technical details):

- Architecture components:
  - Type of the model – for instance, RNN-based, SAN-based (Vaswani et al., 2017), feed-forward-based, Transformers-based (Devlin et al., 2018).
  - Type and size of the layers – including individual layers, such as CNN (Albawi, Mohammed, & Al-Zawi, 2017), LSTM (Hochreiter et al., 1997), GRU (Cho et al., 2014), ResNet (He, Zhang, Ren, & Sun, 2016), AlexNet (Krizhevsky, Sutskever, & Hinton, 2012), and multi-layer architectures, such as BERT (Devlin et al., 2018). These can be applied with or without bidirectionality (Schuster et al., 1997), attention (Bahdanau, Cho, & Bengio, 2015), skip-connections (Chang, Zhang, Han, Yu, Guo, Tan, & Huang, 2017), and other architectural components.
  - Type of input - DNN input is a vector (a series of numbers). Each number can represent a word using word-embedding methods, such as Word2Vec (Mikolov et al., 2013) or GloVe (Pennington et al., 2014), a single character using one-hot encoding or character-embedding (Char2Vec), encoded features, or contextual embeddings (e.g., BERT (Devlin et al., 2018), RoBERTa (Liu, Ott, Goyal, Du, Joshi, Chen, & Stoyanov, 2019), XLNet (Yang, Dai, Yang, Carbonell, Salakhutdinov, & Le, 2019)) based on the surrounding words.

- Number of layers and the DNN information flow – for instance, encoder-decoder architecture (Cho et al., 2014).
- Activation functions (the neuron's function) – including Sigmoid, Tan-h, ReLU, and Softmax.
- Loss functions (the "size" of the training error) – regression tasks can be: i) mean squared error (MSE), ii) mean squared logarithmic error, iii) mean absolute error; for binary classification tasks: i) binary cross-entropy, ii) hinge, iii) squared hinge; for multi-class classification: i) multi-class cross-entropy, ii) sparse multi-class cross-entropy, iii) Kullback- Leibler divergence.

- Hyper-parameters:

- Type and size of the regulation layers – regulation layers reduce overfitting by adding constraints to the DNN. These constraints, such as dropout (Srivastava et al., 2014), L1, and L2, prevent the model from learning the training data and force it to learn the patterns in the data.
- Batch size - the number of examples to use in a single training pass.
- Number of epochs and the epochs' size - the number of iterations on the training data and the number of examples to use during the entire training process.
- Learning rate, method, and configuration - such as stochastic gradient descent (SGD), adaptive moment estimation (Adam) (Kingma & Ba, 2014), and Adagrad (Duchi, Hazan & Singer, 2011).

Theoretically, architecture components are also hyper-parameters. However, from a practical perspective, once architecture components are chosen, they are usually fixed. There are techniques that can be applied to find and set these architecture components and hyper-parameters automatically. These techniques are called AutoML and are suitable for many different DNN models (and classical ML models). However, AutoML has its limitations: it is often costly (training the model repeatedly), does not fit large-scale problems, and may lead to overfitting (Feurer & Hutter, 2019). It is advisable to check AutoML optimization methods such as submodular optimization (Jin, Yan, Fu, Jiang, & Zhang, 2016), grid search (Montgomery, 2017), Bayesian optimization (Melis, Dyer, & Blunsom, 2017), neural architecture search (So, Liang, & Le, 2019), and others (Feurer et al., 2019) or, if the researcher has a hypothesis or intuition about the problem, it is also possible to test multiple architecture components and hyper-parameters combinations manually. Moreover, training a large DNN language model such as a BERT-based model with standard pre-defined hyper-parameters on public cloud servers costs $2,074-$12,571, depending on the hyper-parameters and the corpus size (Devlin et al., 2018), while using neural architecture search (So et al., 2019) to train a DNN language model with hyper-parameters optimized for the specified task costs $44,055–$3,201,722 (Strubell, Ganesh, & McCallum, 2019). Therefore, the budget is another consideration for using some AutoML methods.

Numerous DH studies have demonstrated the importance and the impact of hyper-parameters optimization on the DNN accuracy. Tanasescu, Kesarwani, & Inkpen (2018) optimized hyper-parameters for poetic metaphor classification. They experimented with different activation functions (ReLU, Tan-h for the inner layers and Softmax and Sigmoid for the output layer), number of layers (1-4), number of neurons in each layer (6-306), dropout rate(0-0.9), number of epochs (20-1000), and batch size (20-200). The optimization increased the metaphor classification F-score by 2.9 (from 80.4

to 83.3) and precision by 5.6 (from 69.8 to 75.4). Wang et al. (2016), used a DNN model for Chinese song iambics generation and tested several architecture components. In their research, Wang et al. (2016) added an attention layer (Bahdanau et al., 2015) on top of bidirectional LSTM layers and tested several domain-specific training methods. This DNN domain optimization made it possible to achieve near-human performance. These use-cases emphasize how important it is for DH/LIS experts to understand architecture components and hyper-parameters and their usage.

Domain-specific Dataset Adaptation for DNN

Using DNN models in some domains can also require adaptation of the data (preprocessing) prior to inputting it into the DNN model. A use-case study of Won, Murrieta-Flores, & Martins (2018) aimed to perform Named Entity Recognition (NER) on two historical corpora, Mary Hamilton Papers (modern English from 1750 to 1820) and the Samuel Hartlib collection (early modern English from 1600 to 1660). NER is an NLP task which outputs identification of entity types in text. Entity types can be places, people, or organization names and other "known names". The historical corpus selected in Won et al. (2018) was OCRed and preserved in hierarchical XML files with texts and metadata. DNN models (and the tools used in the study) for NER are not designed to work directly on XML since XML is a graph-based format, and NER is a sequence-based task. It should be noted that there are graph-based DNN models (e.g., Scarselli, Gori, Tsoi, Hagenbuchner, & Monfardini, 2008), but they are not suitable for the NER task. Therefore, Won et al. (2018) needed to adapt their domain data by "translating" the XML markup into text sequences that a DNN model can receive as input. In this preprocessing phase, the researchers took into account the metadata that exists in the domain that was embedded in the XML file, such as authorship, dates, information about the transliteration project, corrections and suggestions made by the transliterators, and particular words and phrases annotated within the body text. Moreover, the square brackets (and their content) added by the transcribers were semi-automatically removed from the text. The metadata was added to the text sequence as labels for the training data to improve the accuracy of the results. Won et al. (2018) did not use DNN models directly but rather used “off the shelf” software to conduct their research. However, they concluded the research with the recognition that using pre-made tools is not sufficient - “*Finally, it must be noted that although this research accomplished the evaluation of the performance of these NER tools, further research is needed to deeply understand how the underlying models work with historical corpora and how they differ.*”

DNN Pipeline Adaptation

DNN models are designed to work in a certain pipeline of components to solve a specific task. For example, a "naïve" DNN based pipeline for the OCR of a book collection will be: 1) scan a book page, 2) use the image as an input to an image-to-text DNN model, 3) use the obtained text or post-process it to correct errors. However, in some cases, it is advisable to design a new domain-specific pipeline to solve the task or increase the model's accuracy. A use-case of such a domain-specific OCR pipeline is presented by Cilia et al. (2020). The goal of the study was identification of the page’s writer for each page of the given medieval manuscript. Medieval handwritten manuscripts present two unique challenges for OCR: 1) first section letters or titles may be drawn as a picture over several lines, and 2) handwritten lines are not always aligned and may reduce accuracy when performing a full-page OCR. Cilia et al. (2020) designed a pipeline for processing handwritten medieval texts with three main steps, using: 1) an object detector to detect lines in the page's scanned image and separate a picture at the top from the text lines, 2) a separate DNN classifier to classify each line, and 3) a majority vote among multiple DNN classifiers obtained for each line and picture object at the line-level, in order to make a decision for the classification (writer identification) of the entire page. This pipeline, tailored to the medieval paleography domain, solved the domain's unique challenges by separating between picture objects and text lines and classifying each line with a different classifier instead of classifying an entire page with a single DNN model (the naïve pipeline). This pipeline’s domain adaptation approach combined with the transfer learning approach, described in the previous section, produced an impressive 96% accuracy in identifying writers that would not have been achieved without this adaptation.

Pipeline adaptation is not just pipelining different models or combining ML and DL; it is also re-training and adapting an existing model, i.e., fine-tuning a model. Fine-tuning a model is a subset of transfer learning, in which a model is trained on a different dataset and also changed by setting different hyper-parameters or adding new last layers on top of the model to fit a specific task. In their research, Todorov and Colavizza (2020), fine-tuned a BERT-based model (Devlin et al., 2018) for increasing the annotation accuracy of NER in French and German historical corpora. In particular, the Groningen Meaning Bank's Corpus Annotated for NER was applied (Bos, Basile, Evang, Venhuizen, & Bjerva, 2017). To embed words (including sub-words) and characters, four models were applied: (1) newly trained word-embeddings on their historical corpus, (2) in-domain pre-trained embeddings that were trained on another corpus in the same domain, (3) BERT-based embedding that was trained on French and German Wikipedia, and (4) character level embeddings learned from the historical corpus training data. As can be observed from Figure 2, Todorov et al. (2020) combined the embedding (by concatenation) and transferred the unified embeddings to a new layer based on a Bi-LSTM-CRF layer. A Bi-LSTM-CRF layer is a Bidirectional (Schuster et al., 1997) Long Short-Term Memory (Hochreiter et al., 1997) layer that merges the sub-word embedding input into a word-level output and transfers its output to fully connected layers (one layer per each entity type) which then outputs tag (entity type) probabilities for each token using Conditional Random Fields (Lafferty, McCallum, & Pereira, 2001). The Bi-LSTM-CRF method has been shown as useful and accurate by Lample, Ballesteros, Subramanian, Kawakami, & Dyer (2016). They also changed the LSTM activation function (remove the tan-h function) and tried three different hyper-parameters configurations. Using the domain-specific pipeline, model, and hyper-parameters, the researchers dramatically increase the accuracy (in some entity types by over 20%) of NER task on French and German historical corpora compared to a state-of-the-art baseline model. Moreover, they tested the impact of the pre-trained generic embedding. They found that (1) without using the open-domain embedding (BERT), their model did not attain high accuracy, and (2) on the other hand, "freezing" the open-domain embedding layers (i.e., using them but re-training only the top layers on the domain-specific historical data) did not affect the accuracy. These findings demonstrate the importance of adapting DNN models to a specific domain and task, while reducing the training time and costs by freezing the large open-domain layers. It is essential to note that besides inputting the historical corpora documents into the DNN model, Todorov et al. (2020) also tested the addition of manually-created features to the documents such as title, numeric and other markups; these features did not have any effect on the accuracy, proving that the DNN model "learned" (or at least did not need) these features.

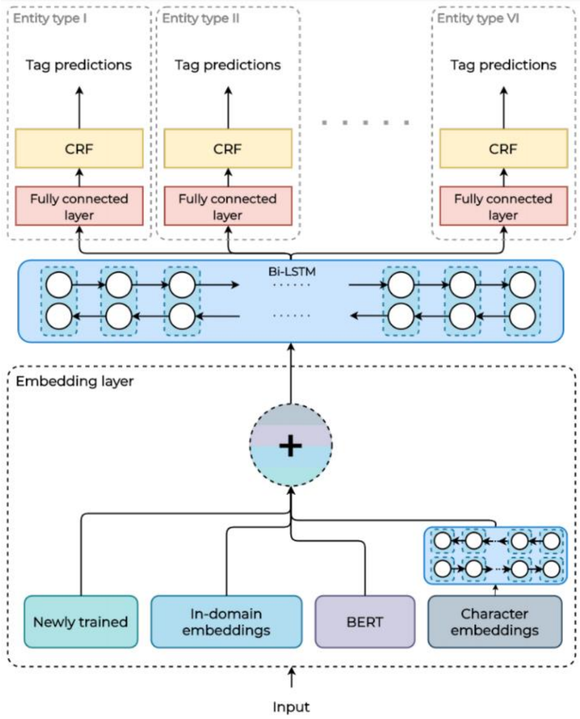


Figure 2: Historical corpora NER fine-tuning pipeline (Todorov & Colavizza, 2020).

A Decision Model for Using Deep Learning for Digital Humanities Research

Based on the above analysis of challenges and possible solutions illustrated by multiple use-case studies described in the recent literature, it is clear that the DH/LIS experts must know just enough math, understand the inner-working of ML and DL algorithms, Python programming, and use these frameworks and other popular modules (Géron, 2019).

Therefore, this paper argues that DH/LIS researchers can no longer see NLP and ML researchers as their "tool makers", and must learn to apply and adapt deep learning models (DNNs) to their specific research domain. However, since working with DNN models requires significant effort, computational resources, budget, and time, a decision model was formulated for assisting DH experts in determining when it is "worthwhile" to invest in training DNN models. The decision model is based on two strategies: 1) the data availability strategy – how to assess the types of methods and models suitable for the available dataset, and 2) the domain adaptation strategy – how to determine whether and when it is “worthwhile” to invest in domain adaptation.

Figure 3 presents the data availability strategy and leads to three possible recommendations: (1) with no data, either zero-shot DL models, or hard-coded rules/assumptions regarding domain data should be implemented, based on prior knowledge and experience; (2) with limited data, either classical machine learning algorithms, such as SVM or HMM, or few-shot DL models can be used; otherwise (3) it is advisable to use supervised deep learning models for the task. It should be noted that if the DNN model is overfitting (high accuracy on the training dataset and low accuracy on the validation dataset), it is advisable to increase the dataset size by employing expert workers, crowdsourcing, or synthetic data generation. Figure 4 presents the domain adaptation strategy and also leads to three possible recommendations: (1) if strict rules can be defined, there is no need for ML or DL; (2) with limited resources or for low accuracy tasks, ML is the preferable option, and (3) with the appropriate resources and a need for high accuracy, DL with domain adaptation should be utilized. A researcher can use both strategies of the proposed decision model to choose the recommended approach for the given task. Since there are many different text analysis tasks, some aspects of the strategies depend on the expert's assessment; for example, "what is considered a small or a large dataset?" and "what is low or high accuracy?". These assessments should be performed by the researcher based on the concrete task, domain, and needs. Notice that the advice to use DNN models does not mean that it is not recommended to combine them with ML algorithms when suitable.

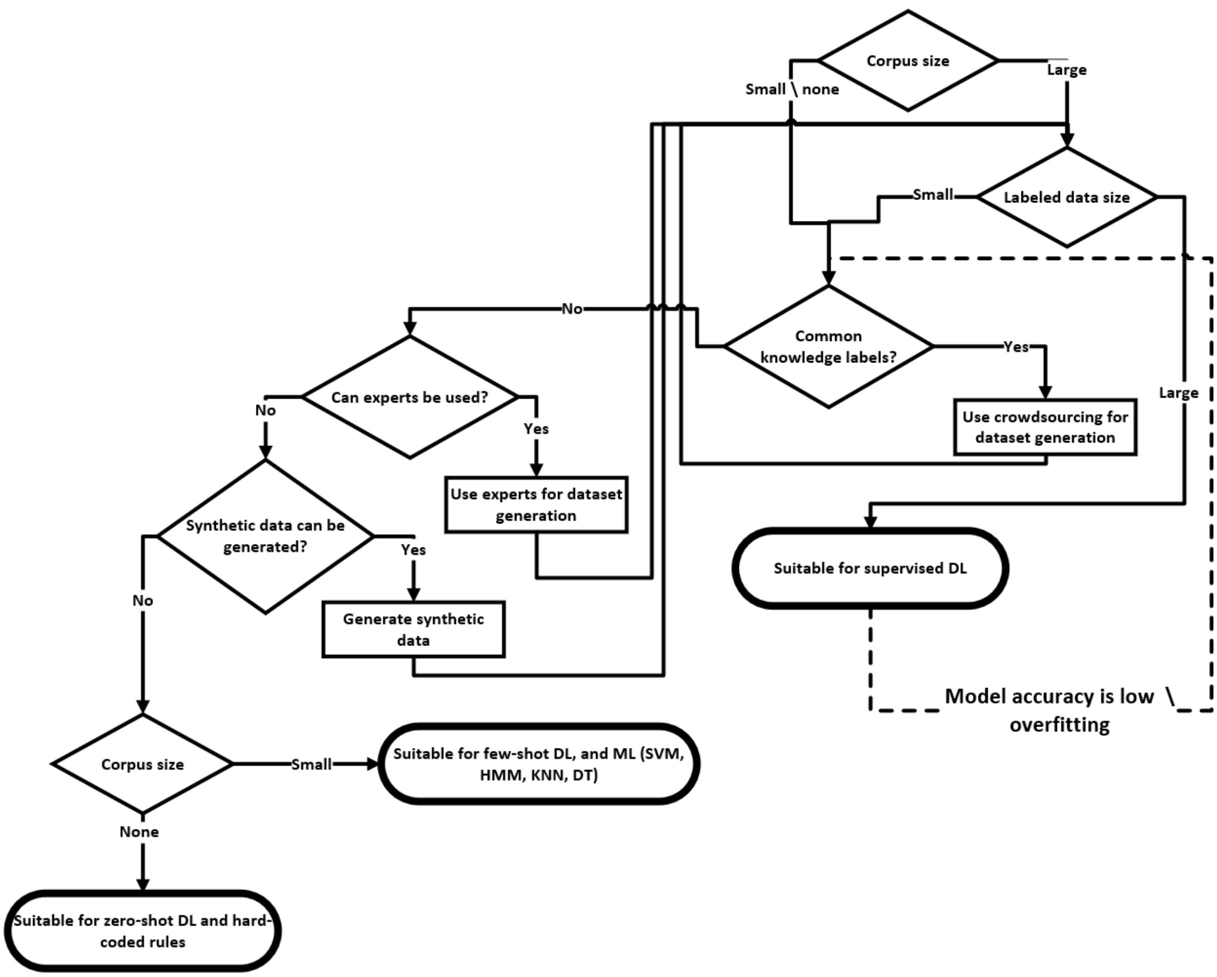


Figure 3: Data availability strategy for DH researchers

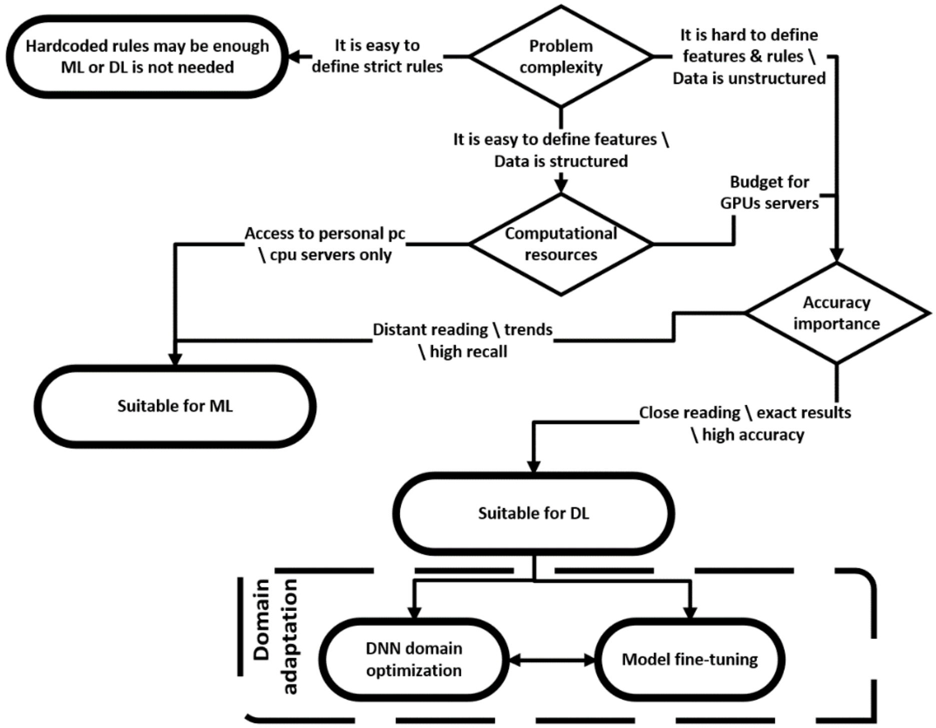


Figure 4: Domain adaptation strategy for DH researchers

As can be observed from the proposed decision model, supervised DL should be used when there is a large corpus (or a large corpus can be generated), for complex problems such as unstructured texts, when the researcher has a budget for computational resources (GPUs servers), and accuracy is essential (domain adaption is always assumed). Since most of the DH corpora are not labeled, dataset generation will most probably be required. When the labeling requires only "common knowledge", it is advisable to use crowdsourcing (if possible); otherwise, the researcher should consider using domain experts or automatic generating of synthetic data as explained above in this paper. A step-by-step example for decision model usage for a specific DH task can be found in Appendix II.

It should be noted that the extensive computational resources needed to train DNN models have an impact on the environment. DL may become a major contributor to climate change if the exponential growth of training more and more DNN models continues (Anthony, Kanding, & Selvan, 2020; Hsueh, 2020). It has been estimated that training one transformer model such as BERT-based (Devlin et al., 2018) will produce similar amounts of CO_2 to those of air travel of one person from NY to SF; using a neural architecture search (So et al., 2019), an AutoML method, will produce almost five times more CO_2 than an average car produces throughout its lifetime including the fuel (Strubell et al., 2019). We note that the proposed decision model does not consider environmental impact, yet researchers should be aware of this and take it into consideration.

By using this decision model as a guideline and applying the suggested solutions for the two fundamental challenges faced by many DH projects – DH-specific training dataset generation and model adaptation, DH/LIS experts can solve a variety of important tasks in the field for diverse national languages, such as 1) improving OCR post-correction (including restoring damaged text); 2) automated ontology and knowledge graph construction for various DH domains (based on entity/category and relation extraction and NER); and 3) corpus-based stylometric analysis and profiling of DH resources (e.g., identification of an author, date, location, and sentiment of the given text or image).

Conclusion and Discussion

This paper presents the main two challenges almost every DH/LIS research can expect to encounter using DNN models in her research. Although classic learning techniques based on rules, patterns, or predefined features are no longer considered state-of-the-art in many text processing tasks (e.g., Thyaharajan, Sampath, Durairaj, & Krishnamoorthy, 2020; Glazkova, 2020), DH/LIS researchers are still using them often, even when there is a better alternative such as deep neural networks. The reasons for avoiding using deep learning in DH may be the lack of "off-the-shelf" tools tailored for the specified task, lack of training data, as well as time, computational resources, and budget limitations. Based on the presented investigation of the main challenges of using DNN in DH research and the proposed decision model for handling these challenges, and the potential adoption of DNN methods, this paper argues that DH/LIS researchers should expand their arsenal of computational skills and methods. A DH expert must acquire in-depth knowledge in mathematics, software programming and have a deep understanding of the usage of deep neural network frameworks. Therefore, we encourage DH/LIS academic departments to introduce the following topics into their academic syllabus, at the applied (rather than theoretical) level:

- Multivariable calculus (partial derivatives, gradients, high order derivatives),
- Linear algebra (vector space, matrices operations, matrices decompositions),
- Probability (distribution, entropy),
- Statistics (bayesian, parameter estimation, overfitting, and underfitting),
- Mathematical optimization (gradient descent, stochastic gradient descent),
- Unsupervised machine learning (k-means, hierarchical clustering, local outlier factor),
- Supervised machine learning (SVM, logistic regression, naïve bayes, knn),
- Unsupervised and self-supervised deep learning (autoencoders, deep belief networks, generative adversarial networks, embeddings),
- Supervised deep learning (feed-forward, RNN, Self-Attention Network (SAN), CNN),
- Python / R programming (working with data, visualization, ML and DL frameworks, working with GPUs).

Adding these topics to the academic syllabus of DH/LIS experts does not mean that DH/LIS experts will become Computer Science experts, but rather they will be able to comprehend and adapt DL algorithms for their needs. Using this knowledge, DH/LIS experts will no longer be limited to "off the shelf" tools developed for generic open-domain tasks, and will be able to utilize the full potential of the DL algorithms.

Finally, in addition to raising awareness of digital humanities researchers of deep neural networks as the state-of-the-art text analysis method, researchers should be encouraged to generate and release public DH/LIS corpora for training deep neural networks.

REFERENCES:

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Kudlur, M. (2016). Tensorflow: A system for large-scale machine learning. In 12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16) (pp. 265-283).

Aker, A., El-Haj, M., Albakour, M.D. and Kruschwitz, U. (2012), "Assessing Crowdsourcing Quality through Objective Tasks", in LREC , pp. 1456-1461.

Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017, August). Understanding of a convolutional neural network. In 2017 International Conference on Engineering and Technology (ICET) (pp. 1-6). IEEE.

Alias-i, L. (2008). 4.1. 0. URL http://alias-i. com/lingpipe.

Almeida, T., Hidalgo, J. M. G., & Silva, T. P. (2013). Towards sms spam filtering: Results under a new dataset. International Journal of Information Security Science, 2(1), 1-18.

Anthony, L. F. W., Kanding, B., & Selvan, R. (2020). Carbontracker: Tracking and predicting the carbon footprint of training deep learning models. arXiv preprint arXiv:2007.03051.

Bahdanau, D., Cho, K. H., & Bengio, Y. (2015, January). Neural machine translation by jointly learning to align and translate. In 3rd International Conference on Learning Representations, ICLR 2015.

Banar, N., Lasaracina, K., Daelemans, W., & Kestemont, M. (2020). Transfer Learning for Digital Heritage Collections: Comparing Neural Machine Translation at the Subword-level and Character-level. In ICAART (1) (pp. 522-529).

Beaudoin, J., & Buchanan, S. (2012). Digital humanities and information visualization: Innovation and integration. Bulletin of the American Society for Information Science and Technology, 38(4), 14-15.

Bengio, Y. (2012). Practical recommendations for gradient-based training of deep architectures. In Neural networks: Tricks of the trade (pp. 437-478). Springer, Berlin, Heidelberg.

Bengio, Y., Goodfellow, I., & Courville, A. (2017). Deep learning (Vol. 1). Massachusetts, USA:: MIT press.

Bergstra, J., Breuleux, O., Bastien, F., Lamblin, P., Pascanu, R., Desjardins, G., ... & Bengio, Y. (2010, June). Theano: a CPU and GPU math expression compiler. In Proceedings of the Python for scientific computing conference (SciPy) (Vol. 4, No. 3, pp. 1-7).

Bernstein, M.S., Little, G., Miller, R.C., Hartmann, B., Ackerman, M.S., Karger, D.R., Crowell, D. and Panovich, K. (2010), "October. Soylent: a word processor with a crowd inside" In Proceedings of the 23nd annual ACM symposium on User interface software and technology. ACM. pp. 313-322.

Berry, D. M. (2011). The computational turn: Thinking about the Digital Humanities. Culture Machine, 12, 1–22.

Biemann, C., Crane, G. R., Fellbaum, C. D., & Mehler, A. (2014). Computational humanities-bridging the gap between computer science and digital humanities (Dagstuhl Seminar 14301). In Dagstuhl reports (Vol. 4, No. 7). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Blanke, T., Bryant, M., & Hedges, M. (2020). Understanding memories of the Holocaust—A new approach to neural networks in the digital humanities. Digital Scholarship in the Humanities, 35(1), 17-33.

Bos, J., Basile, V., Evang, K., Venhuizen, N. J., & Bjerva, J. (2017). The groningen meaning bank. In Handbook of linguistic annotation (pp. 463-496). Springer, Dordrecht.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language Models are Few-Shot Learners. arXiv preprint arXiv:2005.14165.

Bunescu, R., & Mooney, R. (2007, June). Learning to extract relations from the web using minimal supervision. In Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics (pp. 576-583).

Burgoyne, J. A., Fujinaga, I., & Downie, J. S. (2015). Music information retrieval. A new companion to digital humanities, 213-228.

Chang, S., Zhang, Y., Han, W., Yu, M., Guo, X., Tan, W., ... & Huang, T. S. (2017). Dilated Recurrent Neural Networks. Advances in Neural Information Processing Systems, 30, 77-87.

Chen, C. M., & Chang, C. (2019). A Chinese ancient book digital humanities research platform to support digital humanities research. The Electronic Library, 37(2), 314-336.

Chiron, G., Doucet, A., Coustaty, M., & Moreux, J. P. (2017, November). Icdar2017 competition on post-ocr text correction. In 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR) (Vol. 1, pp. 1423-1428). IEEE.

Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014, October). Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 1724-1734). Chowdhary, K. R. (2020). Natural language processing. In Fundamentals of Artificial Intelligence (pp. 603-649). Springer, New Delhi.

Cilia, N. D., De Stefano, C., Fontanella, F., Marrocco, C., Molinara, M., & Freca, A. S. D. (2020). An Experimental Comparison between Deep Learning and Classical Machine Learning Approaches for Writer Identification in Medieval Documents. Journal of Imaging, 6(9), 89.

Ciobanu, A. M., Dinu, L. P., Şulea, O. M., Dinu, A., & Niculae, V. (2013, September). Temporal text classification for romanian novels set in the past. In Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013 (pp. 136-140).

Collobert, R., & Weston, J. (2008, July). A unified architecture for natural language processing: Deep neural networks with multitask learning. In Proceedings of the 25th international conference on Machine learning (pp. 160-167).

Davis, P. T., Elson, D. K., & Klavans, J. L. (2003, May). Methods for precise named entity matching in digital collections. In 2003 Joint Conference on Digital Libraries, 2003. Proceedings. (pp. 125-127). IEEE.

De Smet, H. (2005). A corpus of Late Modern English texts. Icame Journal, 29(29), 69-82.

Deng, L. & Liu, Y. (Eds.). (2018). Deep learning in natural language processing. Springer.

Dereza, O. (2018, October). Lemmatization for Ancient Languages: Rules or Neural Networks?. In

Conference on Artificial Intelligence and Natural Language (pp. 35-47). Springer, Cham.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Downs, J.S., Holbrook, M.B., Sheng, S. and Cranor, L.F. (2010), "' Are your participants gaming the system?: screening mechanical turk workers". In Proceedings of the SIGCHI conference on human factors in computing systems, ACM. pp. 2399-2402.

Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. Journal of machine learning research, 12(7).

Drucker, J., Kim, D., Salehian, I., Bushong, A. (2014). Introduction to the Digital Humanities. Concepts, Methods and Tutorials for Students and Instructors. Los Angeles: University of California Los Angeles. Available at http://dh101.humanities.ucla.edu.

Eiteljorg, H. (2004). Computing for archaeologists. A Companion to Digital Humanities, 20-30.

Elmalech, A., & Dishi, Y. (2021). Cost-Effective Method for Generating Training Data to Machine Learning Models, iConference 2021 Proceedings.

Elmalech, A., & Grosz, B. (2017). " But you Promised": Methods to Improve Crowd Engagement In Non-Ground Truth Tasks. In Proceedings of the AAAI Conference on Human Computation and Crowdsourcing (Vol. 5, No. 1).

Elson, D. K., Dames, N., & McKeown, K. R. (2010). Extracting social networks from literary fiction. In Proceedings of the 48th annual meeting of the Association for Computational linguistics, ACL'10 (pp. 138–147). Stroudsburg, PA: Association for Computational Linguistics.

Ess, C. (2004). Revolution? What Revolution? Successes and Limits of Computing Technologies in Philosophy and Religion. Companion to Digital Humanities. Blackwell Companions to Literature and Culture. Malden, MA and Oxford: Blackwell Publishing, 132-144.

Feurer, M., & Hutter, F. (2019). Hyperparameter optimization. In Automated Machine Learning (pp. 3-33). Springer, Cham.

Finegold, M., Otis, J., Shalizi, C., Shore, D., Wang, L., & Warren, C. (2016). Six degrees of Francis Bacon: a statistical method for reconstructing large historical social networks. Digital Humanities Quarterly, 10(3).

Finkel, J. R., Grenager, T., & Manning, C. D. (2005, June). Incorporating non-local information into information extraction systems by gibbs sampling. In Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05) (pp. 363-370).

Forte, M. (2015). Cyberarchaeology: a Post-Virtual Perspective. Humanities and the Digital. A Visioning Statement. MIT Press. Boston.

Géron, A. (2019). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. O'Reilly Media.

Glazkova, A. (2020). A Comparison of Synthetic Oversampling Methods for Multi-class Text Classification. arXiv preprint arXiv:2008.04636.

Gold, M. K. (Ed.). (2012). Debates in the digital humanities. U of Minnesota Press.

Gold, M. K., & Klein, L. F. (Eds.). (2016). Debates in the Digital Humanities 2016. U of Minnesota Press.

Gulli, A., & Pal, S. (2017). Deep learning with Keras. Packt Publishing Ltd.

Habernal, I., & Gurevych, I. (2017). Argumentation mining in user-generated web discourse. Computational Linguistics, 43(1), 125-179.

Hajic, J. (2004). Linguistics meets exact sciences (Vol. 42, p. 79). Wiley-Blackwell.

Hall, M(2020). Opportunities and Risks in Digital Humanities Research. In: Carius, Hendrikje; Prell, Martin and Smolarski, René eds. Kooperationen in den digitalen Geisteswissenschaften gestalten. Vandenhoeck & Ruprecht GmbH & Co. KG, Göttingen, pp. 47–66.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

Hellrich, J., & Hahn, U. (2016, December). Bad company—neighborhoods in neural embedding spaces considered harmful. In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers (pp. 2785-2796).

Hinrichs, E., Hinrichs, M., Kübler, S., & Trippel, T. (2019). Language technology for digital humanities: introduction to the special issue.

Hochreiter, S., & Schmidhuber, J. J. (1997). Long short-term memory. Neural Computation, 9(8), 1–32. https://doi.org/10.1162/neco.1997.9.8.1735

Hutchings, T. (2015), "Digital humanities and the study of religion", in Svensson, P. and Goldberg, D.T.(Eds), Between Humanities and the Digital, MIT Press, Cambridge, MA, pp. 283-294.

Hsueh, G. (2020). Carbon Footprint of Machine Learning Algorithms. Senior Projects Spring 2020. https://digitalcommons.bard.edu/senproj_s2020/296

Jin, J., Yan, Z., Fu, K., Jiang, N., & Zhang, C. (2016). Neural network architecture optimization through submodularity and supermodularity. stat, 1050, 1.

Kingma, D. P., & Ba, J. (2015, January). Adam: A Method for Stochastic Optimization. In ICLR (Poster).

Kittur, A., Smus, B., Khamkar, S. and Kraut, R.E. (2011), "Crowdforge: Crowdsourcing complex work", In Proceedings of the 24th annual ACM symposium on User interface software and technology, ACM, pp. 43-52.

Kollatz, T. (2019). 18 EPIDAT–Research Platform for Jewish Epigraphy. In Crossing Experiences in Digital Epigraphy (pp. 231-239). De Gruyter Open Poland.

Koltay, T. (2016), "Library and information science and the digital humanities: perceived and real strengths and weaknesses", Journal of Documentation, Vol. 72 No. 4.

Krapivin, M., Autaeu, A., & Marchese, M. (2009). Large dataset for keyphrases extraction. University of Trento.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25, 1097-1105.

Kuhn, J. (2019). Computational text analysis within the Humanities: How to combine working practices from the contributing fields?. Language Resources and Evaluation, 53(4), 565-602.

Kumar, A., & Lehal, G. S. (2016). Automatic text correction for Devanagari OCR. Indian Journal of Science and Technology, 9(45).

Lafferty, J., McCallum, A., & Pereira, F. C. (2001). Conditional random fields: Probabilistic models for

segmenting and labeling sequence data.
Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., & Dyer, C. (2016). Neural architectures for named entity recognition. arXiv preprint arXiv:1603.01360.
Li, J., Sun, A., Han, J., & Li, C. (2020). A Survey on Deep Learning for Named Entity Recognition. IEEE Computer Architecture Letters, (01), 1-1.
Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... & Zitnick, C. L. (2014, September). Microsoft coco: Common objects in context. In European conference on computer vision (pp. 740-755). Springer, Cham.
Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.
Liu, Z., Lv, X., Liu, K., & Shi, S. (2010, March). Study on SVM compared with the other text classification methods. In 2010 Second international workshop on education technology and computer science (Vol. 1, pp. 219-222). IEEE.
Maas, A., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., & Potts, C. (2011, June). Learning word vectors for sentiment analysis. In Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies (pp. 142-150).
McCarty, W. (2013, July). Becoming interdisciplinary. In DH (pp. 293-295).
McGillivray, B., Poibeau, T., & Ruiz Fabo, P. (2020). Digital Humanities and Natural Language Processing:"Je t'aime... Moi non plus".
Melis, G., Dyer, C., & Blunsom, P. (2017). On the state of the art of evaluation in neural language models. arXiv preprint arXiv:1707.05589.
Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems (pp. 3111-3119).
Mintz, M., Bills, S., Snow, R., & Jurafsky, D. (2009, August). Distant supervision for relation extraction without labeled data. In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP (pp. 1003-1011).
Montgomery, D. C. (2017). Design and analysis of experiments. John wiley & sons.
Moreno-Ortiz, A. (2017, April). Lingmotif: Sentiment analysis for the digital humanities. In Proceedings of the Software Demonstrations of the 15th Conference of the European Chapter of the Association for Computational Linguistics (pp. 73-76).
Needleman, S. B., & Wunsch, C. D. (1970). A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins. Journal of Molecular Biology, 48(3), 443–453.
Neudecker, C. (2016, May). An open corpus for named entity recognition in historic newspapers. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16) (pp. 4348-4352).
Niculae, V., Zampieri, M., Dinu, L. P., & Ciobanu, A. M. (2014, April). Temporal text ranking and automatic dating of texts. In Proceedings of the 14th Conference of the European Chapter of the

Association for Computational Linguistics, volume 2: Short Papers (pp. 17-21).

Pantel, P., & Pennacchiotti, M. (2006, July). Espresso: Leveraging generic patterns for automatically harvesting semantic relations. In Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics (pp. 113-120).

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... & Desmaison, A. (2019). Pytorch: An imperative style, high-performance deep learning library. In Advances in neural information processing systems (pp. 8026-8037).

Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1532-1543).

Poole, A. H. (2017). The conceptual ecology of digital humanities. Journal of Documentation.

Posner, M. (2013). No half measures: Overcoming common challenges to doing digital humanities in the library. Journal of Library Administration, 53(1), 43-52.

Prebor, G., Zhitomirsky-Geffet, M., Buchel, O., & Bouhnik, D. (2018). A New Methodology for Error Detection and Data Completion in a Large Historical Catalogue Based on an Event Ontology and Network Analysis. Digital Humanities 2018: Book of Abstracts/Libro de resúmenes.

Prebor, G., Zhitomirsky-Geffet, M., & Miller, Y. (2020a). A new analytic framework for prediction of migration patterns and locations of historical manuscripts based on their script types. Digital Scholarship in the Humanities, 35(2), 441-458.

Prebor, G., Zhitomirsky-Geffet, M., & Miller, Y. (2020b). A Multi-dimensional Ontology-based Analysis of the Censorship of Hebrew Manuscripts. *Digital Humanities Quarterly, 14(1*).

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI blog, 1(8), 9.

Rajpurkar, P., Zhang, J., Lopyrev, K., & Liang, P. (2016). Squad: 100,000+ questions for machine comprehension of text. arXiv preprint arXiv:1606.05250.

Rajpurkar, P., Jia, R., & Liang, P. (2018). Know what you don't know: Unanswerable questions for SQuAD. arXiv preprint arXiv:1806.03822.

Ramsay, S. (2016). Who's in and who's out. In Defining digital humanities (pp. 255-258). Routledge.

Rigaud, C., Doucet, A., Coustaty, M., & Moreux, J. P. (2019, September). ICDAR 2019 competition on post-OCR text correction. In 2019 International Conference on Document Analysis and Recognition (ICDAR) (pp. 1588-1593). IEEE.

Robinson, L., Priego, E., & Bawden, D. (2015). Library and information science and digital humanities: two disciplines, joint future?. Re-inventing information science in the networked society.

Rokach, L. and Maimon, O. (2015) *Data mining with decision trees: theory and applications*. 2nd editio. Hackensack New Jersey: World Scientific.

Rommel, T. (2004). Literary studies. In A Companion to Digital Humanities, ed. Susan Schreibman, Ray Siemens, John Unsworth. Oxford: Blackwell. http://www.digitalhumanities.org/companion/

Rossum, G. (1995). Python reference manual.

Rubinstein, A. (2019). Historical corpora meet the digital humanities: the Jerusalem Corpus of

Emergent Modern Hebrew. Language Resources and Evaluation, 53(4), 807-835.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323, 533–536. 159, 330

Russell, I. G. (2011). The role of libraries in Digital Humanities. Erişim adresi: http://www.ifla.org/past-wlic/2011/104-russell-en. pdf.

Saltz, D. Z. (2004). Performing arts. A Companion to Digital Humanities, 121-131.

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2008). The graph neural network model. IEEE Transactions on Neural Networks, 20(1), 61-80.

Schulz, S., & Ketschik, N. (2019). From 0 to 10 million annotated words: part-of-speech tagging for Middle High German. Language Resources and Evaluation, 53(4), 837-863.

Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing, 45(11), 2673–2681. https://doi.org/10.1109/78.650093

Singhal, A. (2001). Modern information retrieval: A brief overview. IEEE Data Eng. Bull., 24(4), 35-43.

Skadiņš, R., Tiedemann, J., Rozis, R., & Deksne, D. (2014, May). Billions of parallel words for free: Building and using the eu bookshop corpus. In Proceedings of LREC.

Snow, R., O'Connor, B., Jurafsky, D. and Ng, A.Y. (2008), "Cheap and fast---but is it good?: evaluating non-expert annotations for natural language tasks", In Proceedings of the conference on empirical methods in natural language processing, Association for Computational Linguistics, pp. 254-263.

So, D. R., Liang, C., & Le, Q. V. (2019). The evolved transformer. arXiv preprint arXiv:1901.11117.

Sorokin, A. and Forsyth, D. (2008), "Utility data annotation with amazon mechanical turk". In 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, IEEE, pp. 1-8.

Spinaci, Gianmarco, Colavizza, Giovanni, & Peroni. (2019). List of Digital Humanities journals (Version 0.1.0) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.3406564.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research, 15, 1929–1958. https://doi.org/10.1214/12-AOS1000

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. arXiv preprint arXiv:1906.02243.

Suissa, O., Elmalech, A., & Zhitomirsky-Geffet, M. (2019). Toward the optimized crowdsourcing strategy for OCR post-correction. Aslib Journal of Information Management, *72(2),* 179-197.

Suissa, O., Elmalech, A., & Zhitomirsky-Geffet, M. (2020). Optimizing the neural network training for OCR error correction of historical Hebrew texts. iConference 2020 Proceedings.

Sula, C. A. (2012). Visualizing social connections in the humanities: Beyond bibliometrics. Bulletin of the American Society for Information Science and Technology, 38(4), 31-35.

Sula, C.A. (2013), "Digital humanities and libraries: a conceptual model", Journal of Library Administration, Vol. 53 No. 1, pp. 10-26.

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to Sequence Learning with Neural Networks. International Journal of Computer Vision.

Svensson, P. (2010). The landscape of digital humanities. Digital Humanities, 4(1).
Tanasescu, C., Kesarwani, V., & Inkpen, D. (2018, May). Metaphor detection by deep learning and the place of poetic metaphor in digital humanities. In The Thirty-First International Flairs Conference.
Thomas, W. G. (2004). Computing and the historical imagination. A companion to digital humanities, 56-68.
Thompson, P. (2017). The voice of the past: Oral history. Oxford university press.
Thyaharajan, S. K., Sampath, K., Durairaj, T., & Krishnamoorthy, R. SSN NLP at CheckThat! 2020: Tweet Check Worthiness Using Transformers, Convolutional Neural Networks and Support Vector Machines.
Todorov, K., & Colavizza, G. (2020, January). Transfer Learning for Named Entity Recognition in Historical Corpora. In CLEF (Working Notes).
Torrey, L., & Shavlik, J. (2010). Transfer learning. In Handbook of research on machine learning applications and trends: algorithms, methods, and techniques (pp. 242-264). IGI global.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In Advances in neural information processing systems (pp. 5998-6008).
Vijayan, V. K., Bindu, K. R., & Parameswaran, L. (2017, September). A comprehensive study of text classification algorithms. In 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI) (pp. 1109-1113). IEEE.
Wang, Q., Luo, T., Wang, D., & Xing, C. (2016, July). Chinese song iambics generation with neural attention-based model. In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (pp. 2943-2949).
Warwick, C. (2012). Institutional models for digital humanities. Digital humanities in practice, 193-216.
Williams, R. R., Hunt, S. C., Heiss, G., Province, M. A., Bensen, J. T., Higgins, M., ... & Hopkins, P. N. (2001). Usefulness of cardiovascular family history data for population-based preventive medicine and medical research (the Health Family Tree Study and the NHLBI Family Heart Study). The American journal of cardiology, 87(2), 129-135.
Won, M., Murrieta-Flores, P., & Martins, B. (2018). ensemble named entity recognition (ner): evaluating ner Tools in the identification of Place names in historical corpora. Frontiers in Digital Humanities, 5, 2.
Wooldridge, R. (2004). Lexicography. A Companion to Digital Humanities, 69-78.
Yang, E., Ravikumar, P. K., Allen, G. I., & Liu, Z. (2013). On Poisson graphical models. Advances in Neural Information Processing Systems, 26, 1718-1726.
Yang, S., Wang, Y., & Chu, X. (2020). A Survey of Deep Learning Techniques for Neural Machine Translation. arXiv preprint arXiv:2002.07526.
Yang, Y., Bansal, N., Dakka, W., Ipeirotis, P., Koudas, N. and Papadias, D. (2009), "Query by document", In Proceedings of the Second ACM International Conference on Web Search and Data Mining, ACM, pp. 34-43.
Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., & Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. arXiv preprint arXiv:1906.08237.
Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning based natural

language processing. ieee Computational intelligenCe magazine, 13(3), 55-75.

Zaagsma, G. (2013). On digital history. BMGN-Low Countries Historical Review, 128(4), 3-29.

Zampieri, M., & Becker, M. (2013). Colonia: Corpus of historical portuguese. ZSM Studien, Special Volume on Non-Standard Data Sources in Corpus-Based Research, 5, 69-76.

Zhitomirsky-Geffet, M., & Prebor, G. (2019). SageBook: Toward a cross-generational social network for the Jewish sages' prosopography. Digital Scholarship in the Humanities, 34(3), 676-695.

Zhitomirsky-Geffet, M., Prebor, G. & Miller, I. (2020). Ontology-based analysis of the large collection of historical Hebrew manuscripts. Digital Scholarship in the Humanities, 35(3), 688-719.

Zhou, G., & Su, J. (2002, July). Named entity recognition using an HMM-based chunk tagger. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (pp. 473-480).