

Extracting Causal Claims from Information Systems Papers with Natural Language Processing for Theory Ontology Learning

Roland M. Mueller
Berlin School of Economics and Law
roland.mueller@hwr-berlin.de

Sebastian Huettemann
Berlin School of Economics and Law
sebastian.huettemann@gmail.com

Abstract

The number of scientific papers published each year is growing exponentially. How can computational tools support scientists to better understand and process this data? This paper presents a software-prototype that automatically extracts causes, effects, signs, moderators, mediators, conditions, and interaction signs from propositions and hypotheses of full-text scientific papers. This prototype uses natural language processing methods and a set of linguistic rules for causal information extraction. The prototype is evaluated on a manually annotated corpus of 270 Information Systems papers containing 723 hypotheses and propositions from the AIS basket of eight. F1-results for the detection and extraction of different causal variables range between 0.71 and 0.90. The presented automatic causal theory extraction allows for the analysis of scientific papers based on a theory ontology and therefore contributes to the creation and comparison of inter-nomological networks.

Keywords: Causal Relationship Extraction, Theory Ontology Learning, Natural Language Processing, Theory, Causality

1. Introduction

There is an exponential increase in the number of scientific papers published each year [37] containing countless theories and causal claims. The behavior of humans, teams, and organizations that are affected by Information Systems is therefore described by hundreds of different theories in various disciplines [12, 31]. These theories often have minor or major variations and are evaluated and tested in different settings. Understanding and evaluating these inter-nomological networks [33] is becoming more time-consuming and difficult—especially if adjacent disciplines shall be included in the analysis [29].

Causal propositions and hypotheses are an important part of many Information Systems theories [17, 18, 38]. Combining, interlinking and building

upon these theories is a major aspect of scientific progress. However, a researcher who wants to find, for example, all the influencing factors for the success of an e-learning initiative or the potential causes for knowledge sharing in a knowledge management system must still rely on a manual search in order to find connections, because he or she is only supported by existing literature databases with full-text search and co-citation analysis.

An automated procedure that detects and extracts hypotheses and propositions as the main theoretical content from scientific publications would support the combination and interlinking of theories. Therefore, at least two things are necessary: (1) a structured description of theoretical claims in the form of a theory ontology or meta-model [38, 39], which helps isolating the elements (e.g. cause and effect) of a theory and making it possible to connect them with complementary research from other publications and (2) a software-prototype that is able to automatically detect and extract the relevant contents from scientific publications into the given structure.

In this paper, a software-prototype is introduced that integrates with the meta-model of Mueller [38, 39] and is able to detect and extract hypotheses and propositions from scientific papers. Because theoretical propositions in Information Systems and other fields like management or social sciences often have not just causes and effects, but also moderators, mediators, conditions, and signs (causal directions), this prototype is developed for and tested on extracting causal theories out of papers belonging to these domains. To the best of the authors' knowledge, this combination of causal extraction tasks has not yet been addressed.

The remaining part of the paper is structured as follows: Section 2 describes the problem of causal extraction, defines theoretical causal claims, and discusses ontologies for causal theories. Section 3 presents related work regarding rule-based and machine-learning based approaches for causal relationship extraction. Section 4 describes CauseMiner, a software-prototype that applies a rule-based natural language processing approach. Section 5

contains the evaluation of this prototype followed by the conclusion and ideas for future work.

2. Problem Description

Gregor [17:620] distinguishes between different means of theory representation: (a) words, (b) mathematical terms, (c) symbolic logic, (d) diagrams, (e) tables, or (f) graphically. This paper focuses on the extraction from natural language (a: words). This extracted information should then be translated and represented based on a theory ontology or theory meta-model [38]. Gregor [17] did not mention theory ontologies or formal meta-models for theories. In her classification of theory representations, it is closest to (c) symbolic logic. From a formal theory representation, diagrams (d) can be generated [38]. However, not just visualization but also causal reasoning is supported by a theory ontology [38, 39]. We define the main causal theoretical claims of a paper as all sentences that (a) are part of the theory presented in the paper and (b) have at least implicitly a causal meaning (which might be causally ambiguous).

Not all causal sentences are main causal claims and therefore relevant for further analysis. Excluded are causal sentences that state other causal claims that are unrelated to the theory, for example, that a specific circumstance of the analyzed organization affected the choice of research methods or that the relevance of the research question caused the author to analyze the research problem. In Information Systems papers, main causal claims are typically described as hypotheses or propositions. Therefore, the proposed software prototype only analyzes and extracts hypotheses and propositions.

Some hypotheses, as shown in Figure 1, do not explicitly state a causal relationship between constructs but just a relationship, even if the constructs are (causally) linked in the structure equation model (SEM). According to Pearl [41] an SEM should be interpreted causally. Therefore, we also include causally ambiguous hypotheses and rather err on including too many than too few hypotheses in our analyzes.

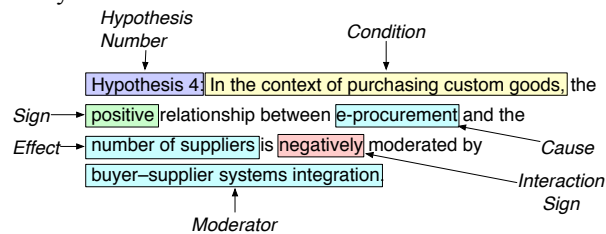


Figure 1. Example of a Hypothesis [13]

The goal of CauseMiner is the extraction of a causal theory's structure from a paper. This structure is based on the meta-model of causal theories of Mueller [38, 39]. Figure 1 shows an example of a hypothesis from an Information Systems paper with different parts that should be extracted [13].

The extraction of theoretical structures from natural language is difficult because of (a) the detection of hypothesis and proposition sentences within the paper, (b) the numerous different possible expressions in natural language of the same logical causal structure, (c) an unclear description of the theory, (d) an incomplete description caused by e.g. omitting the scope of theory or using acronyms for variable names (e) the use of non-causal language for causal relationships, and (f) the use of different names for the same construct or the same name for different constructs [30]. Also, (g) scientific papers are often imprecise and their description of theories is sometimes incomplete and ambiguous [49]. Therefore, causal extraction with text-mining from natural language is a challenging task.

For each hypothesis in a scientific paper, CauseMiner tries to extract the following parts: hypothesis number, hypothesis, cause, effect, sign (causal direction), moderator, mediator, interaction sign, and condition. These parts are consistent with the description of theory parts by Gregor [17], Weber [49], and Mueller [38] and supported by an analysis of example hypotheses in Information Systems papers.

Cause and effect are constructs that are linked by a causal proposition or hypothesis. A typical construct in Information Systems research is e.g. 'perceived ease of use'. The sign describes the relation between cause and effect and can have the following values: +, -, ?, 0 [38]. A positive sign (+) means increasing the value of the cause makes a higher value of the effect more likely even if all other constructs are hold constant. A negative sign (-) is to interpret vice versa. Question marks (?) indicate an ambiguous influence whereas zeroes (0) stand for no direct causal effect at all [38]. Some hypotheses explicitly mention that there is no direct causal effect (sign 0) between two constructs. This is implicitly assumed between all constructs in a theory that are not directly linked with a hypothesis [38]. A moderator is a variable that affects the direction and/or strength of the relation between an independent and a dependent variable [38]. The mediator is a construct that explains the relation between an independent and a dependent variable. Moderator variables indicate when specific effects will hold [8], i.e. an additional variable that influences the strength of the relation between cause and effect. Mediator variables explain how or why these effects take place [8], i.e. the causal chain between cause and

effect goes through the mediator-variable. Conditions refer to a theory's scope or boundary. These statements of boundaries show the limits of generalization [17], e.g. only under a certain condition, there is a relationship between two variables.

3. Related Work

Research in causal relationship extraction can broadly be divided into two categories: rule-based methods and machine learning methods [6]. Rule-based methods refer to the application of pre-defined rules to natural language text. These rules can consist of encoded grammatical structures that incorporate causal keywords, e.g. *<NP1 verb NP2>* [15]. Thereby, they enable the detection of explicitly or implicitly stated causal relationships. Machine learning methods refer to the application of computer-based statistical learning techniques. One example for this is to train a classifier with manually annotated text examples of causal relationships. Based on this training, a classifier might be able to identify underlying rules by applying different statistical methods [6].

Li & Larsen [33] used a rule-based system for detecting sentences as hypotheses and machine learning for construct (i.e. variable) extraction in Information Systems papers. For the rule-based hypotheses detection task, they achieved 85.39% precision, 84.64% recall, and 84.99% f-measure. For the machine learning approach of variable extraction, they achieved 74.65% precision, 69.59% recall, 72.02% f-measure, and 89.92% accuracy. They used the extracted constructs for theory integration. Because the constructs are not labeled as cause and effect, their approach is not directly considered as causal relationship extraction.

The first publications on causal information extraction are from Khoo et. al [24, 26, 27], who worked on detecting explicitly expressed causal relationships in natural language text. They developed an automatic method based on linguistic cues for identifying and extracting cause-effect information from Wall Street Journal text. This method provided an extensive collection of linguistic patterns designed to identify causal relationships within a sentence and also in adjacent sentences [24]. Khoo et. al [27] focused on the domain-independent aspect of causal relationship extraction and classified different ways to express causal relationships in English text:

1. *Causal links* are typical words or phrases that link cause and effect, e.g. 'hence', 'therefore', 'because of', 'on account of', etc. The different classifications of causal links were based on Altenberg [3].
2. *Causal verbs* are verbs that mean to cause something, be caused by something, prevent something

from happening, or affect something, e.g. 'lead to', 'trigger', 'enable' (based on Thomson [48]).

3. *Resultative constructions* refer to sentences where the object of a verb is followed by a phrase that describes the state of the object as a result of the action implied by the verb, e.g. 'I painted the car yellow.' These constructions are based on Simpson [43].

4. *Conditionals* often indicate a causal relationship where the antecedent (the *if* part) causes the consequent (the *then* part) (based on [11]).

5. *Causal adverbs and adjectives* often have a causal element in their meaning, e.g. 'Brutus fatally wounded Caesar' (based on [10]).

Girju and Moldovan [15] presented an approach for the automatic detection of causal patterns and semi-automatic validation of ambiguous lexico-syntactic patterns referring to causation. One of the results is a list of causal verbs containing words like 'generate', 'associate', 'originate', etc. ordered by ambiguity and frequency which is often used by other researchers. This approach was novel insofar as linguistic patterns were extracted semi-automatically and no longer by hand as it was done in previous research. The accuracy on a corpus of news articles was 0.65.

Sorgente et. al [44] worked on techniques that incorporated different variations of machine learning approaches. They designed a hybrid system relying on rules and a Bayesian classifier. Precision and recall were between 0.6 and 0.7.

Ackermann [1] compared among other aspects different machine learning techniques on a set of New York Times news articles. F-measures were 0.74 for Naive Bayes, 0.78 for dictionary classifiers, 0.58 for KNN, 0.67 for SVM, 0.70 for decision trees and 0.71 for neural networks.

Yang and Mao [50] as well as Pakray and Gelbukh [40] developed additional machine learning systems achieving F-measures between 0.65 and 0.85 for the detection of causal sentences.

In the related work, rule-based papers used pattern matching, linguistic clues, cue phrases, and causal verbs as methods for causal information extraction [6]. The main machine learning methods applied were decision trees, SVM, and Naive Bayes [6].

Most corpora in the related work were derived from relatively broad sources like newspapers or Wikipedia. Most of the research regarding causal relationship extraction focuses on domain-independent approaches for general corpora [6]. Also, the prior research mainly checks for causality but rarely addresses the extraction of causal variables beyond cause and effect, e.g. moderator and mediator variables as well as causal directions between constructs.

4. CauseMiner

CauseMiner is a software that analyzes text in order to identify and extract sentences containing causal relationships. It detects, analyzes, and extracts hypotheses and propositions from research papers. CauseMiner is developed and tested on papers that belong to the Senior Scholars' Basket of Journals [7]. The decision for these papers is based on the relevance for the Information Systems research community and the domain expertise of the authors.

CauseMiner extracts a corpus of TXT-files (after conversion from PDF to text) and saves results into a CSV-file. It applies a rule-based approach and is written in Python. Rule-based approaches are still quite effective for some complex information extraction tasks without a lot of annotated data [6, 20, 34] and could also be combined with machine learning approaches within hybrid systems [6, 20].

4.1. Pre-Processing

Figure 2 provides an abstract overview of the pre-processing steps. The analyzed papers are available as PDF and converted into text format with the open source software pdftotext [16]. This conversion of PDF files into text causes significant problems regarding data quality. After conversion, text includes page breaks or header/footer information that would impede further processing. An analysis of conversion errors revealed variations of possible data problems for each journal. Figure 3 provides two examples of conversion errors that interrupt the text-flow.

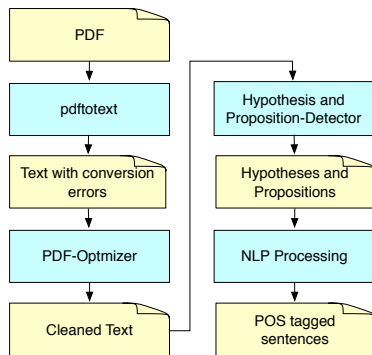


Figure 2. Preprocessing

Because the name of a journal isn't necessarily part of the filename nor otherwise indicated, CauseMiner detects the name of a journal per regular expressions and executes specific clean up procedures, e.g. removal of header/footer information, concatenation of page breaks, removal of references, figures, etc. This is achieved by a set of rules for each journal, e.g. *if the*

actual line is empty, check if seven lines later only 1-4 numbers occur. If yes, remove the lines in-between. These rules can be easily adjusted for other journals that are not part of the AIS basket of eight.

European Journal of Information Systems		
Theoretically, these five phases can be iterated indefinitely as every iterative cycle is supposed to generate 'further knowledge about the organization and the validity of relevant theoretical frameworks' (Baskerville, 1999, p 17). However, for pragmatic reasons, the data collection stage concludes		
European Journal of Information Systems	Interfering Header/ Footer- Information	
Managing user acceptance towards ERP		
Eric T.K. Lim et al		
139	Interfering Footnote	
when information appears to have reached saturation, that is, further iterations do not lead to additional insights about the research problem at hand (Glaser & Strauss, 1967).		
MIS Quarterly		
In this paper, we draw on the symbolic consumption literature (e.g., Grubb and Grathwohl 1967) to theorize how environmental stimuli that facilitate		
2"Active VW User Forecast: 2009–2013," KZER Worldwide (http://www.kzero.co.uk/blog/?p=3836#more-3836). 3In March 2010 alone, 201,707 residents in Second Life spent between L\$1 and L\$500; 17,128 residents spent between L\$50,000 and L\$100,000, and 2,581 spent over L\$500,000 (i.e., US\$2,000) (http://secondlife.com/statistics/economy-data.php).		
social interactions in virtual worlds might shape purchase behaviors (Arnett et al. 2003; Simon 2004).		

Figure 3. PDF-conversion errors [4, 35]

The analysis of embedded hypotheses in text data reveals many variations. Each journal has a specific style of expressing hypotheses and propositions that can additionally vary depending on the author's preferences. Explicitly expressed hypotheses and propositions are almost always preceded by an indicator, e.g. "H1:" or "P2a:". However, variations are manifold: "Hypothesis 3 (H3).", "P3A.", "Proposition 3.1:", "H1a-e:", "P1", etc. CauseMiner detects hypotheses and propositions after the text cleaning procedure via regular expression. Because of the many possible variations of different indicators, the design of a robust regular expression for this task is difficult.

Therefore, the implemented regex is kind of over-tolerant in order to catch all possible variations. The implemented regex looks like this:

```
"((([^\w]H|^\wP|Hypothesis H|Hypothesis|Proposition P|Proposition)\s?([0-9]+[.]?[0-9]*[aA-zZ])?( \(.?*\))?[.:]*[s]+(.?*[a-z]{0,}[.])?W*?)"
```

However, catching all occurrences of different types of hypotheses statements in a paper is

problematic especially when the regex catches not only the initial hypotheses but also snippets from the result section, e.g. "H1a: $p < 0.01$ ". To address this over-tolerance, additional logic is applied where the prototype tests, if e.g. the extracted contents are numbers or letters. Only valid hypotheses and propositions are processed further.

Parsing detected hypotheses and propositions allows for the application of more sophisticated rules, e.g. filtering for words and word types. CauseMiner utilizes SpaCy [21], a Python library for natural language processing that is fast and has an accuracy within 1% of the best available NLP libraries [9, 21, 22]. Each hypothesis and proposition is parsed into separate word objects and each of these objects is enriched with index-, string-, lemma-, and POS-information. Figure 4 illustrates the parsing of a hypothesis.

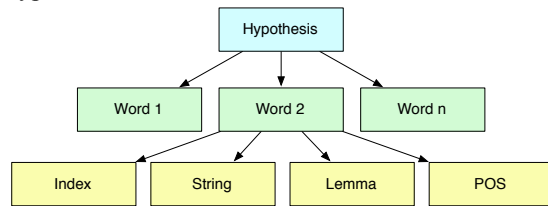


Figure 4. Hypothesis parsing

4.2. Information Extraction Process

An analysis of extracted hypotheses and propositions showed that conditions typically appear at the beginning or end of a sentence. Therefore, they are extracted first. The extracted part is cut off from the sentence and only remaining words are processed further. Moderator and mediator variables are extracted next and again, only the remaining words are analyzed further. Lastly, cause and effect are extracted. This extraction order is the result of experimentations with different rule variations and architecture designs. This is in line with design as a search process according to Hevner [19]. Starting with the extraction of causal patterns resulted for example in complications with extracting conditions later on, because each extracted variable had to be checked separately. Cutting already extracted parts off the sentence during the process prevented redundancy in the code and proved to be more robust. Figure 5 illustrates the extraction process.

For extracting variables as separated constructs, several extraction-patterns were developed. The goal was to extract hypotheses and propositions into the meta-model of Mueller [38] consisting of cause, effect, sign, mediator, moderator, interaction sign, and condition. Therefore, the authors manually extracted 274 papers of the basket of eight and copied all detected 963 hypotheses and propositions into a

spreadsheet (see Section 5.1 for more details on the corpora). For every hypothesis and proposition, all variables (cause, effect, etc.) were stored separately into designated columns in the same spreadsheet.

This manual extraction revealed different types of causal statements that were clustered based on the results of former research in this area. The resulting categories or patterns were the basis for the development of extraction procedures within the prototype.

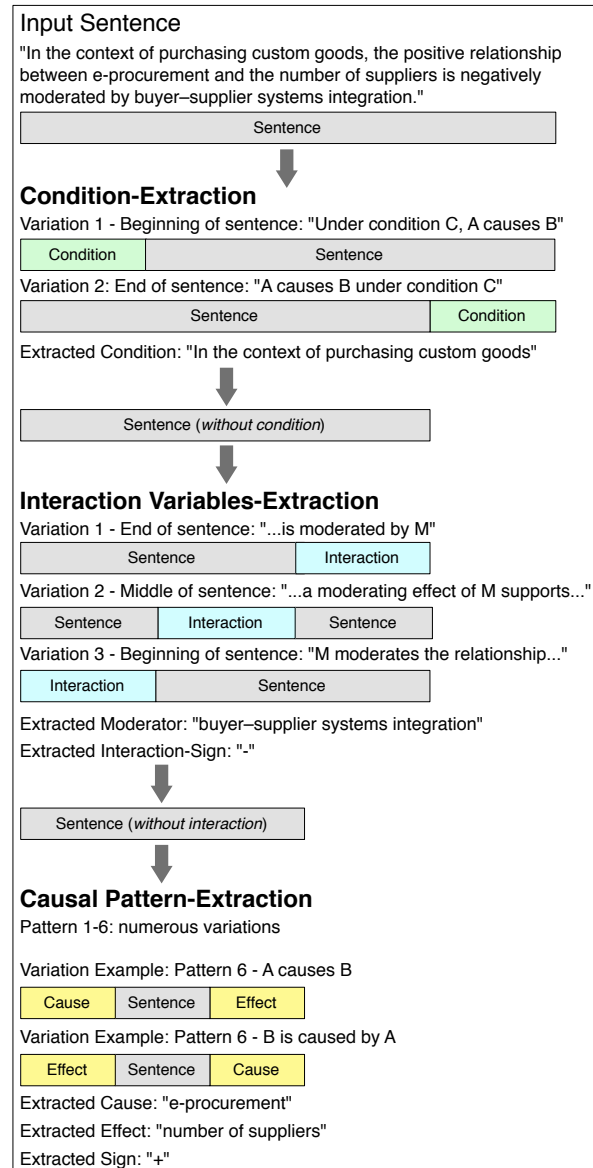


Figure 5. Hypothesis extraction

Table 1 provides an overview of the extraction patterns. The example pattern column contains an abstract example of the underlying rules: brackets hold examples of keywords, C_i stands for construct, rel for

Table 1. Causal extraction patterns

Pattern Name	Example Pattern	Example Sentence	Based on
Causal Relationships	[relationship, interaction, linkage, etc. between] C ₁ and C ₂	"H2: MTA positively moderates the relationship between perceived usefulness (adoption) and intention to use." [46]	
Causal Keywords	C ₁ [because, as, since, etc.] C ₂	"H2a: When using audio, unilateral concessions will increase due to the communication of fact-related questions." [23]	[3, 25]
Causal Comparisons	C ₁ rel C ₂ [greater than] C ₁ rel C ₃	"Hypothesis 5b (H5b). Firms that use advanced BA will not generate higher positive abnormal returns than firms that use basic BA." [47]	[42]
Causal Correlations	[the more, higher, faster] C ₁ , [the lower, fewer, less] C ₂	"Hypothesis 3: The higher an employee's resistance to change, the higher her user resistance behaviour." [32]	[42]
Causal Verbs	C ₁ [causes, affects, increases] C ₂	"Hypothesis 2: Cohesive argument structuring relates positively to deception severity in CMC." [36]	[15, 48]
Conditions	If Cond., C ₁ [causes] C ₂ ; If Cond., rel between C ₁ and C ₂	"Hypothesis 2 (H2A). In contexts in which initial technology trust expectations are more mature, positive trust disconfirmation will negatively influence trusting intention." [28]	
Moderator, Mediator	M [moderates, mediates] rel between C ₁ and C ₂	"H2a. The relationship between the level of e-government development in a country and the extent of corruption in its national business service systems is mediated by the corruption in its national political institutions." [45]	

relation. Each pattern includes numerous variations, e.g. distinction between active and passive sentences, sign detection (positive, negative, not mentioned, negated) or detection of patterns in different parts of a sentence (beginning, middle, end).

The example patterns shown in Table 1 are just small excerpts that illustrate the general rule-logic. Each pattern includes many variations: the pattern "Causal Verbs" includes, for example, grammatical variations such as "C₁ causes/caused/will cause/is going to cause/etc. C₂" as well as "C₂ is caused by/has been caused by/was caused by/etc. C₁". It also includes different keywords such as "impacts", "influences", "increases", "stimulates", etc.

After extraction, CauseMiner removes stop words from constructs and saves results in table-form in a CSV-file. The CSV-files (or alternatively a different table-format or a database) are used to run automated tests that compare annotated CSV-files with the results of the prototype. It also allows further processing of single variables in later development stages that might be covered in a future version of CauseMiner (e.g. matching of synonyms).

The fieldnames in this file are "Hypothesis-Number", "Hypothesis/Proposition", "Pattern-Number", "Cause", "Effect", "Sign", "Condition", "Moderator", "Mediator", "Interaction-Sign", "Journal", "Author", "Year", "Title", and some binary control variables. For example, the hypothesis "MTA positively moderates the relationship between perceived usefulness (adoption) and intention to use." [46] would be saved as:

- Hypothesis-Number: '3a'
- Hypothesis: 'MTA positively moderates the relationship between perceived usefulness

(adoption) and intention to use.'

- Pattern: '1'
- Cause: 'perceived usefulness (adoption)'
- Effect: 'intention to use'
- Sign: '?'
- Condition_Flag: 'False'
- Condition: ''
- Interaction_Flag: 'True'
- Moderator: 'MTA'
- Mediator: ''
- Interaction Sign: '+'
- Journal: 'Journal of the Association for Information Systems'
- Author: 'Sun et. al'
- Year: '2016'
- Title: 'Choosing a Fit Technology: Understanding Mindfulness in Technology Adoption and Continuance'

While iterating over a corpus of papers, CauseMiner adds one row for each detected hypothesis/proposition into a CSV-file that can be further analyzed.

5. Evaluation

5.1. Corpus Creation

In order to test the prototype for the Information Systems domain, the authors manually analyzed three corpora and created several goldmaster files. The annotation process consisted of manually analyzing PDF-files and extracting explicitly stated hypotheses and propositions into an Excel-sheet—a goldmaster file. For each hypothesis, the following variables were

placed into separate columns: hypothesis, hypothesis-number, cause, effect, causal sign, moderator, mediator, interaction sign, condition, journal, author, title.

This process was conducted by two annotators (the authors): (1) creation of an annotation protocol, (2) collection of papers that belong to the basket of eight: For each year (2007–2015), the amount of papers should be equally distributed, (3) manual checking of papers for the existence of hypotheses/propositions and the documentation of findings in a CSV file (this step was done once by each annotator): If a hypothesis/proposition was found, all relevant variables were manually extracted in the provided CSV fields, and (4) discussion of deviations regarding the annotators judgments: if manual extraction-results deviated from each other, the annotators agreed on a final desired extraction result.

The first corpus consisted of 124 papers containing 474 hypotheses and propositions. It led to the development and implementation of an initial rule-set. The second corpus consisted of 150 papers containing 489 hypotheses and propositions. This second corpus helped to refine and optimize the initial rules. All papers belonged to journals of the AIS basket of eight [7] and were published from 2008 till 2015. The third corpus was designed to evaluate the improved rule-based approach and was created after the rule-based system was finalized. This corpus contains all hypotheses of all 270 papers published in 2016 in the AIS basket of eight containing 723 hypotheses and propositions.

5.2. Evaluation

The extraction method was tested against the manually created goldmaster file. Table 2 provides an overview of the test corpus from 2016 for the evaluation step. Table 3 shows the distribution of the causal extraction patterns in the test corpus.

Table 2. Goldmaster corpus overview

Detection Results	Total	Rel.
Papers in corpus	270	1.0
Papers without causal claims	154	0.57
Papers containing causal claims	116	0.43

For each variable, true positives, false positives, true negatives, false negatives, precision, recall, F-measure, and accuracy were computed. Table 4 contains the definitions for all measures. Measures for variables (cause, effect, etc.) were only computed for those hypotheses that were detected.

When CauseMiner did not detect a hypothesis that should have been detected, only false positives for the

hypothesis-variable were affected, but not for other variables. Only exact matches are considered true positives. Because one missing word in a construct might drastically change the meaning of the construct, partial matches are considered as false.

Table 3. Causal Extraction Pattern Distribution

Detection Results	Total	Rel.
Extraction patterns were applied	564	0.82
Extraction patterns couldn't be applied	127	0.18
Extracted Pattern	Total	Rel.
Causal Relationships	104	0.18
Causal Keywords	25	0.04
Causal Comparisons	54	0.10
Causal Correlations	61	0.11
Causal Verbs	287	0.51
Other (e.g. condition + effect, no cause)	33	0.06
Additional Pattern	Total	Rel.
Conditions	65	0.09
Moderator	27	0.04
Mediator	9	0.01

Table 4. Measures for evaluation

Measure	Description
True positive (TP)	Variable in goldmaster-file equals variable in computed output and is not empty ('found something and it is correct')
False positive (FP)	Variable in goldmaster-file doesn't equal variable in computed output ('found something but it was the wrong thing')
True negative (TN)	Both variables in goldmaster-file and in computed output are empty ('found nothing and it is correct')
False negative (FN)	Variable in computed output is empty, but variable in goldmaster-file is not ('found nothing although something should have been found')
Precision	$\frac{TP}{TP + FP}$
Recall	$\frac{TP}{TP + FN}$
F-Measure (F1-Score)	$2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$

Table 5 shows the comparison of CauseMiner's output against the annotated goldmaster file. In the hypothesis detection task, CauseMiner achieved 0.88 precision, 0.88 recall, 0.88 F-measure, and 0.82 accuracy. Compared to the hypotheses detection of Li & Larsen [33] (precision 0.85, recall: 0.85, F-Measure: 0.85) these are slightly better results.

The results of the causal relationship extraction of CauseMiner are only to a limited amount comparable

to the related work, because not only cause and effect, but also sign, moderator, mediator, condition, and interaction sign were extracted. Another limitation regarding the comparability of the results is that the related work mainly uses different corpora often from news and medical abstracts. Also, not every paper published all relevant metrics like precision, recall, F-measure, and accuracy but just a subset or only one, like accuracy or precision.

Table 5. Evaluation results

Variable	TP	FP	FN	TN
Hypothesis	616	82	85	148
Cause	401	108	109	26
Effect	408	122	95	19
Sign	513	17	95	19
Condition	60	5	50	529
Moderator	27	0	17	600
Mediator	9	0	12	623
Int.-Sign	36	0	29	579
Variable	Prec.	Recall	F1	Acc.
Hypothesis	0.88	0.88	0.88	0.82
Cause	0.79	0.79	0.79	0.66
Effect	0.77	0.81	0.79	0.66
Sign	0.97	0.84	0.90	0.83
Condition	0.92	0.55	0.69	0.91
Moderator	1.0	0.61	0.76	0.97
Mediator	1.0	0.43	0.60	0.98
Int.-Sign	1.0	0.55	0.71	0.95

5.3. Error Analysis and Limitations

An error analysis showed different categories and sources whereas none of them are regarded as major issues. Nearly all of them could be resolved by optimizing different parts of the prototype. Table 6 provides an overview of some typical errors.

Other problems refer, for example, to multiple variables. The following hypothesis contains actually four hypotheses in one sentence: "H1: For information search via mobile Web browsing, the tree-view presentation adaptation will lead to (a) reduction in search time, (b) increase in search accuracy, (c) increase in perceived ease of use, and (d) increase in perceived usefulness." [2]. This hypothesis contains multiple causal statements and is therefore difficult to extract.

Hypotheses and propositions containing mathematical expressions are problematic as well. Due to the pdf to text conversion, mathematical symbols are converted to Unicode and this results in hieroglyphic statements that can't be processed any further, e.g. "The better i's optimal betting strategy is $r_i^* = E[X | s_i] = \tau_0 s_0 + \tau_1 s_i + \tau_0 + \tau_1 B_i^* = \tau_{i2} \tau_{02} (\tau_0 + \tau_1)^2$ with a positive expected payoff $\tau_{i2} (\tau_0 + \tau_1)^2$." [14].

Some semantic challenges must also be addressed, for example hypotheses that contain statements of comparison: "When using a KBS with explanation facilities, experts will choose more feedback explanations than novices." [5]. Currently, CauseMiner extracts a condition ('When using a KBS with explanation facilities'), a cause ('experts'), an effect ('feedback explanations') and a sign ('+'). But the comparison referring to novices is not extracted.

Table 6. Error examples

Category	Expected	Retrieved
Sentence Boundary Detection	"Perceived effectiveness of WASS on e-commerce transaction intention is stronger for the U.S.A. consumers than for Korean consumers."	"Perceived effectiveness of WASS on e-commerce transaction intention is stronger for the U."
Encoding Errors (wrong apostrophe sign)	"The higher an IT worker's agreeableness, the higher her organizational commitment."	"The higher an IT worker's agreeableness, the higher her organizational commitment."
Not ignoring parenthesis	"For both static and polymorphic warnings, attention will decrease in terms of mouse cursor movements (greater area under the curve, slower average speed, and slower initial acceleration) across repeated exposures."	Extracted cause: "area under the curve"
Truncating errors	Cause: "Paradox-framed trajectories"	Extracted cause: "Paradox-framed trajectories are strongly"

6. Conclusion and Outlook

This paper presents a novel method and a working prototype that automatically extracts not only causes and effects but also signs, moderators, mediators, conditions, and interaction signs from hypotheses and propositions published in scientific papers. Additionally, a review of related literature was conducted that shows that most of the related work uses either news articles or medical papers as a corpus. This paper specifically addresses Information Systems papers of the AIS basket of eight for developing and testing its approach. For the task of theory ontology learning and theory integration this corpus is regarded more relevant than, for example, news articles.

Due to the limited sample size and the nature of the corpus it is hard to make claims regarding the generalizability of the achieved results. However, as the prototype was developed and fine-tuned by developing two corpora consisting of 274 papers containing 963 hypotheses and propositions in total, as well as being tested over a corpus of 270 papers with 723 hypotheses, the results are promising—especially because the prototype proved to be robust enough to catch the majority of expressions in a domain known for rather complex phrasing.

Construct matching was not a part of this paper, but shall be implemented in the future. Larsen and Bong [30] developed a hybrid method for construct detection based on latent semantic analysis (LSA) and knowledge-based methods. In future work, a similar construct matching approach could be used to further analyze the output of CauseMiner.

CauseMiner is sensitive to errors during the pre-processing and extraction steps (e.g. errors in the PDF parsing or part-of-speech tagging) which can lead to wrongly formatted input for additional analytical techniques. Further optimization should make CauseMiner more robust for these kinds of errors.

Asghar [6:8] found no paper that used Deep Neural Networks for causal relationship extraction but suggests it might be effective because of the feature abstraction capabilities. This might be an interesting future approach for CauseMiner.

Furthermore, a hybrid combination of CauseMiner's rule-based approach with machine learning methods might be a promising future possibility.

In the future, CauseMiner will be used to analyze the Information Systems field from the last 20 years including construct matching [30]. CauseMiner could then be used to describe papers based on a theory ontology. This would enable research to analyze and integrate many theories with, for example, theory-evolution graphs [38] or theory-data maps [39].

7. References

- [1] Ackerman, E.J.M. Extracting Causal Relations between News Topics from Distributed Sources. *Dissertation*, Technische Universität Dresden (2013).
- [2] Adipat, B., Zhang, D., and Zhou, L. The effects of tree-view based presentation adaptation on mobile web browsing. *MIS Quarterly* 35, 1 (2011), 99–121.
- [3] Altenberg, B. Causal Linking in Spoken and Written English. *Studia Linguistica* 38, 1 (1984), 20–69.
- [4] Animesh, A., Pinsonneault, A., Yang, S.-B., and Oh, W. An odyssey into virtual worlds: exploring the impacts of technological and spatial environments on intention to purchase virtual products. *MIS Quarterly* 35, 3 (2011), 789–810.
- [5] Arnold, V., Clark, N., Collier, P.A., Leech, S.A., and Sutton, S.G. The differential use and effect of knowledge-based system explanations in novice and expert judgment decisions. *MIS Quarterly* 30, 1 (2006), 79–97.
- [6] Asghar, N. Automatic Extraction of Causal Relations from Natural Language Texts: A Comprehensive Survey. *arXiv preprint arXiv:1605.07895*, (2016).
- [7] Association for Information Systems. Website - Senior Scholars' Basket of Journals. 2017. <https://aisnet.org/SeniorScholarBasket>.
- [8] Baron, R.M. and Kenny, D.A. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology* 51, 6 (1986), 1173.
- [9] Choi, J.D., Tetreault, J.R., and Stent, A. It Depends: Dependency Parser Comparison Using A Web-based Evaluation Tool. *ACL (1)*, (2015), 387–396.
- [10] Cresswell, M.J. Adverbs of Causation. In *Adverbial Modification*. Springer Netherlands, Dordrecht, 1981, 173–192.
- [11] Cummins, D.D., Lubart, T., Alksnis, O., and Rist, R. Conditional reasoning and causation. *Memory & cognition* 19, 3 (1991), 274–282.
- [12] Davis, R., Campbell, R., Hildon, Z., Hobbs, L., and Michie, S. Theories of behaviour and behaviour change across the social and behavioural sciences: a scoping review. *Health Psychology Review* 9, 3 (2015), 323–344.
- [13] Dedrick, J., Xu, S.X., and Zhu, K.X. How Does Information Technology Shape Supply-Chain Structure? Evidence on the Number of Suppliers. *Journal of Management Information Systems* 25, 2 (2008), 41–72.
- [14] Fang, F., Guo, Z., and Whinston, A.B. Collective Outsourcing to Market (COM): A market-based framework for information supply chain outsourcing. *Journal of the Association for Information Systems* 9, 3/4 (2008), 98–118.
- [15] Girju, R. and Moldovan, D. Text mining for causal relations. *Proceedings of the FLAIRS Conference*, (2002), 360–364.
- [16] Glyph & Cog. Website - Xpdf Project. 2017. <http://www.foolabs.com/xpdf/about.html>.
- [17] Gregor, S. The nature of theory in information systems. *MIS quarterly* 30, 3 (2006), 611–642.
- [18] Gregor, S. and Hovorka, D. Causality: The elephant in the room in information systems epistemology. *Proceeding of the European Conference on Information Systems (ECIS)*, (2011).
- [19] Hevner, A.R., March, S.T., Park, J., and Ram, S. Design science in information systems research. *MIS Quarterly* 28, 1 (2004), 75–105.
- [20] Hogenboom, F., Frasincar, F., Kaymak, U., de Jong, F., and Caron, E. A Survey of event extraction methods from text for decision support systems. *Decision Support Systems* 85, (2016), 12–22.
- [21] Honnibal, H. Website - spaCy: Industrial-Strength Natural Language Processing. 2017. <https://spacy.io>.

- [22] Honnibal, M. and Johnson, M. An Improved Non-monotonic Transition System for Dependency Parsing. *EMNLP*, (2015), 1373–1378.
- [23] Johnson, N.A., Cooper, R.B., and Holowczak, R.D. The impact of media on how positive, negative, and neutral communicated affect influence unilateral concessions during negotiations. *European Journal of Information Systems* 25, 5 (2016), 391–410.
- [24] Khoo, C. Automatic Identification of Causal Relations in Text and their use for Improving Precision in Information Retrieval - PhD Thesis - Syracuse University. 1995.
- [25] Khoo, C., Chan, S., and Niu, Y. The many facets of the cause-effect relation. In *The Semantics of Relationships*. Springer, 2002, 51–70.
- [26] Khoo, C.S., Chan, S., and Niu, Y. Extracting causal knowledge from a medical database using graphical patterns. *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics*, (2000), 336–343.
- [27] Khoo, C.S., Kornfilt, J., Oddy, R.N., and Myaeng, S.H. Automatic extraction of cause-effect information from newspaper text without knowledge-based inferencing. *Literary and Linguistic Computing* 13, 4 (1998), 177–186.
- [28] Lankton, N.K., McKnight, D.H., Wright, R.T., and Thatcher, J.B. Using Expectation Disconfirmation Theory and Polynomial Modeling to Understand Trust in Technology. *Information Systems Research* 27, 1 (2016), 197–213.
- [29] Larsen, K., Lee, J., Li, J., and Bong, C. A Transdisciplinary Approach to Construct Search and Integration. *AMCIS 2010 Proceedings*, (2010).
- [30] Larsen, K.R. and Bong, C.H. A tool for addressing construct identity in literature reviews and meta-analyses. *MIS Quarterly* 40, 3 (2016), 1–23.
- [31] Larsen, K.R., Hovorka, D., West, J., et al. Theory Identity: A Machine-Learning Approach. *47th Hawaii International Conference on System Sciences (HICSS)*, IEEE (2014), 4639–4648.
- [32] Laumer, S., Maier, C., Eckhardt, A., and Weitzel, T. User personality and resistance to mandatory information systems in organizations: a theoretical model and empirical test of dispositional resistance to change. *Journal of Information Technology* 31, 1 (2016), 67–82.
- [33] Li, J. and Larsen, K. Establishing Nomological Networks for Behavioral Science: a Natural Language Processing Based Approach. *ICIS 2011 Proceedings*, (2011).
- [34] Lim, C.-G. A survey of temporal information extraction and language independent features. *International Conference on Big Data and Smart Computing (BigComp)*, (2016), 447–449.
- [35] Lim, E.T.K., Pan, S.L., and Tan, C.W. Managing user acceptance towards enterprise resource planning (ERP) systems – understanding the dissonance between user expectations and managerial policies. *European Journal of Information Systems* 14, 2 (2005), 135–149.
- [36] Ludwig, S., van Laer, T., de Ruyter, K., and Friedman, M. Untangling a Web of Lies: Exploring Automated Detection of Deception in Computer-Mediated Communication. *Journal of Management Information Systems* 33, 2 (2016), 511–541.
- [37] Michels, C. and Schmoch, U. The growth of science and database coverage. *Scientometrics* 93, 3 (2012), 831–846.
- [38] Mueller, R.M. A Meta-Model for Inferring Inter-Theory Relationships of Causal Theories. *Proceedings of the 48th Hawaii International Conference on System Sciences (HICSS)*. IEEE Computer Society Press, (2015), 4908–4917.
- [39] Mueller, R.M. Theory-Data Maps: A Meta-Model and Methods for Inferring and Visualizing Relationships between Causal Theories and Empirical Evidences. *Proceedings of the 49th Hawaii International Conference on System Sciences (HICSS)*. IEEE Computer Society Press, (2016), 5288–5297.
- [40] Pakray, P. and Gelbukh, A. An Open-Domain Cause-Effect Relation Detection from Paired Nominals. In A. Gelbukh, F.C. Espinoza and S.N. Galicia-Haro, eds., *Nature-Inspired Computation and Machine Learning*. Springer International Publishing, 2014, 263–271.
- [41] Pearl, J. The causal foundations of structural equation modeling. In *Handbook of Structural Equation Modeling*. Guilford Press, New York, 2012.
- [42] Sevastopoulos, J. Website - Grammar Quizzes, English Grammar Reference: Resource, Practices & Exercises. 2017. <http://www.grammar-quizzes.com>.
- [43] Simpson, J. Resultatives. In L. Levin, M. Rappaport and A. Zaenen, eds., *Papers in Lexical-functional Grammar*. Bloomington: Indiana University Linguistics Club, 1983, 143–157.
- [44] Sorgente, A., Vettigli, G., and Mele, F. Automatic Extraction of Cause-Effect Relations in Natural Language Text. *Proceedings of the 7th International Workshop on Information Filtering and Retrieval - DART@AI*IA 2013*, (2013), 37–48.
- [45] Srivastava, S.C., Teo, T.S., and Devaraj, S. You Can't Bribe a Computer: Dealing with the Societal Challenge of Corruption Through ICT. *MIS Quarterly* 40, 2 (2016), 511–526.
- [46] Sun, H., Fang, Y., and others. Choosing a Fit Technology: Understanding Mindfulness in Technology Adoption and Continuance. *Journal of the Association for Information Systems* 17, 6 (2016), 377–412.
- [47] Teo, T.S.H., Nishant, R., and Koh, P.B.L. Do shareholders favor business analytics announcements? *The Journal of Strategic Information Systems* 25, 4 (2016), 259–276.
- [48] Thomson, J.J. Verbs of action. *Synthese* 72, 1 (1987), 103–122.
- [49] Weber, R. Evaluating and Developing Theories in the Information Systems Discipline. *Journal of the Association for Information Systems* 13, 1 (2012), 1–30.
- [50] Yang, X. and Mao, K. Multi level causal relation identification using extended features. *Expert Systems with Applications* 41, 16 (2014), 7171–7181.