

SteamRoller: replicable and scalable empirical humanities

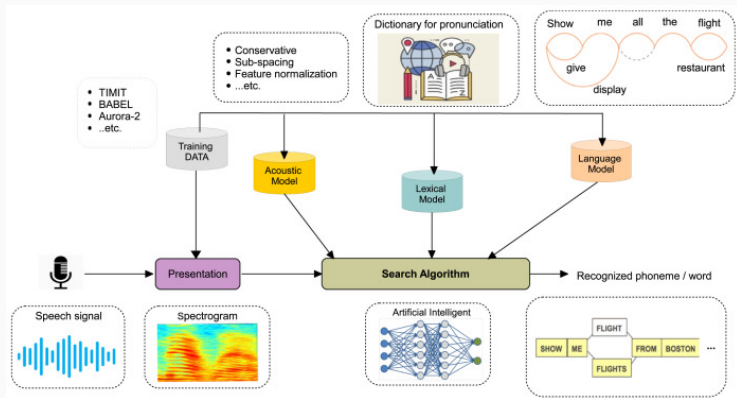
<https://pypi.org/project/steamroller/>



Tom Lippincott

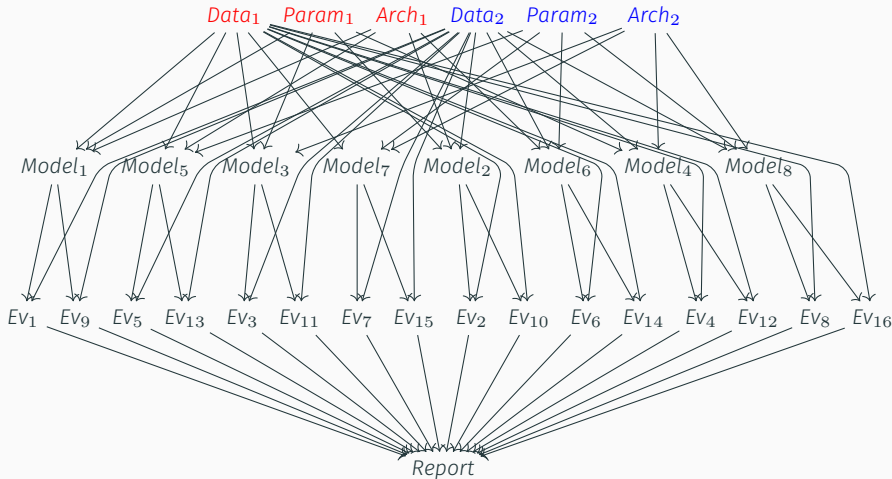
October 7, 2024

Motivation: IARPA BABEL program



- Experiments were (and often still are) *shell scripts*
- So much human error and confusion...

Motivation: Experiments quickly become intractably complex



- Familiar problem in CS ("build systems" like *Make*)
- Goals: Clear, replicable, scalable

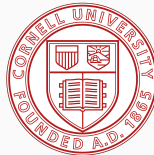
Approach: Augment the SCons build system

- Everything is Python
- Track dependencies
- Map to grid computing (Rockfish and DSAI)

```
TrainModel = Builder(  
    "train_model.py --param ${VALUE} --model_type ${TYPE} --data ${DATA} --model ${MODEL}"  
)  
...  
model = TrainModel(  
    "work/model_${VALUE}_${TYPE}_${DATA}.bin",  
    training_set,  
    VALUE=1.0,  
    TYPE="CNN"  
)
```

- Examining Language Modeling Assumptions Using an Annotated Literary Dialect Corpus
- Detecting Narrative Patterns in Biblical Hebrew and Greek
- How Much Information Do We Need to Transfer Extreme Authorial Style?
- Characterizing the Effects of Translation on Intertextuality using Multilingual Embedding Spaces
- Pairing Orthographically Variant Literary Words to Standard Equivalents Using Neural Edit Distance Models

Impact: Collaboration



- Intro to Computation for the Humanities
- **Experimental Design for the Humanities**
- Intro to Machine Learning for the Humanities

Examples of SteamRoller use from performance period

Detecting Narrative Patterns in Biblical Hebrew and Greek

Hope McGovern¹ Hale Sirin² Tom Lippincott³ Andrew Caines⁴
¹Department of Computer Science & Technology, University of Cambridge, U.K.
²Center for Digital Humanities, Johns Hopkins University, Baltimore, U.S.A.

1.

WE QUANTITATIVELY EXPLORE REPEATING NARRATIVE PATTERNS IN DIGITIZED ANCIENT HEBREW AND GREEK BIBLICAL MANUSCRIPTS.

WE TRAIN AN INFORMATION RETRIEVAL (IR) NETWORK ON A DATASET OF BIBLICAL CROSS-REFERENCES, THEN QUERY THE NETWORK FOR INSTANCES OF NARRATIVE PATTERNS, CALLED "TYPE SCENES".

Type-Scene Description

Type-Scene Description	Example	Scholarly Support
1. A man meets a woman by a well	Genesis 24:11	Ahler (1978)
2. An angel announces a barren woman will bear a child	Judges 13:3	Ahler (1983)
3. The younger son is punished by the older son	Markus 12:3	Dovson (1995)
4. God sends a prophet to speak to a rebellious people	Isaiah 1:1-3	Long (1973)
5. A man has an epiphany in the field	Ezekiel 1:1	Ahler (1978)
6. A prophet visits his flock	Ezekiel 34:33	Betz (2002)
7. A well is found in the desert	Genesis 30:25	Ahler (1978)
8. The words of a dying hero	Judges 16:6	Ahler (1978)

Image 1. 1-8. The images provided are either in the public domain or in the public domain. All rights reserved and no copyright claim. The users should refer to the original source for more information. The images are provided for educational purposes only.

INTRODUCTION

The term type-scene (typische Szene) was coined in the Homeric scholarship to refer to the formulaic and recurring narrative patterns or scenes, such as the arming or sacrifice scenes in the Iliad (Gardt, 1933). This term was later adapted to refer to the phenomenon in Biblical narratives of repeated use of similar patterns to describe a narrative event (Ahler, 1978)

2.

	Ancient Hebrew	Ancient Greek
verses	23,275	31,227
chapters	929	1132
books	39	64
ref pairs	17,899	45,297

OBJECTIVE

We pose type-scene extraction as an information retrieval task, where one verse (i.e. query) and its surrounding context can be used to retrieve thematically similar verses (i.e. documents). Cross-references provide structured information about related passages, whether through direct text reuse, thematic parallelism, or common characters.



ACKNOWLEDGMENTS We thank Juan Antonio Gomez, PhD, professor in the Faculty of Asian Middle Eastern Studies at Cambridge, for his generous help in identifying and collecting the type-scene queries from Biblical Hebrew. This work is supported by the Royal Institute for Cultural Heritage and the Cambridge Trust.

3.

"In the beginning, God created the heavens and the earth." (Gen 1:1)

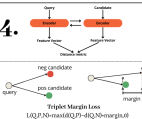
INTUITION

Learning representations that maximize the probability of cross-reference passages may enable us to surface other forms of narrative construction in type-scene.

"In the beginning was the Word, and the word was with God and the Word was God." (John 1:1)



4.



RETRIEVED TYPE-SCENES

In most cases, our network retrieves relevant examples in the top 10 highest ranked candidates.

THE WIDOW AT THE WELL

It happened, late one afternoon, when David arose from his couch and was walking on the roof of the king's house, that he saw from the roof a woman bathing and the woman was very beautiful. (2 Samuel 11:2, Rank 3)

So he arose and went to Jerusalem, And when he came to the gate of the city, behold, a widow was there gathering sticks. And he called to her and said, "Bring me a little water in a vessel, that I may drink." (1 Kings 17:10, Rank 5)

THE HERCULEAN BIRTH

And the angel of the LORD appeared to the woman and said to her, "Behold, you are barren and have no borne children, but you shall conceive and bear a son. (Judges 13:3, Rank 2)

A PROMPT VEILS HIS FACE

He says in his heart, "God has forgotten, he has hidden his face, he will never see it." (Psalm 10:11, Rank 3)

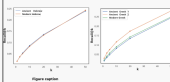
THE WORDS OF A DYING HERO

"My days are number than a razor; they flee away; they see no good. (Job 9:25), Rank 1

5.



7.



CHALLENGES IN INTERPRETATION

- Complexity of data annotation and translation efforts
- Linguistic difficulties in crafting queries
- Confirmation bias

CONCLUSION

Our model may be applied to any long text genre where repeating narrative patterns may be of interest, such as Shakespeare's plays, Greek tragedy, and even movie scripts.

We show that a network fine-tuned on a set of Bible cross-references improves recall over the unsupervised pre-trained models. Although recall values remain low in real terms, our models retrieve relevant candidates in the top 10 results for each type-scene query, observing that this approach holds great promise as a tool for biblical scholars to uncover recurrent narrative patterns.

RELATED LITERATURE

Examples of SteamRoller use from performance period

Princess	Bacteria	Disease	Electricity	Waves
prince	organisms	malady	tricity	billows
empress	microorganisms	fever	electrical	wave
lady	bacilli	infection	trical	ripples
countess	microbes	diseases	elec	wavelets
haiatelnefous	bacterial	infectious	electri	undulations
queen	pneumococci	diphtheria	cyperiments	susurrus
daughter	teria	contagious	pneumaticw	vibrations
king	germs	malaria	magnetism	sounil
bride	streptococci	tuberculosis	treetricity	undampened
betrothed	cocci	influenza	practical	sighings

Examples of SteamRoller use from performance period

