Tanya Goyal

Email: tanyagoyal@utexas.edu
Website: https://tagoyal.github.io/

Education

University of Texas at Austin — Austin, Texas
PhD in Computer Science (Advisor: Greg Durrett) — 2019 – Present
MS in Computer Science — 2017 – Present

Indian Institute of Technology, Guwahati — Guwahati, India
B.Tech in Mathematics and Computing — 2011 – 2015

Select Publications

SNaC: Coherence Error Detection for Narrative Summarization
Tanya Goyal, Jessy Li, Greg Durrett.
EMNLP 2022

HydraSum: Disentangling Style Features in Text Summarization with Multi-Decoder Models
Tanya Goyal, Nazneen Rajani, Wenhao Liu, Wojciech Kryściński.
EMNLP 2022

FALTE: A Toolkit for Fine-grained Annotation for Long Text Evaluation
Tanya Goyal, Jessy Li, Greg Durrett.
EMNLP-Demo 2022

Training and Fine-tuning Dynamics for Summarization
Tanya Goyal, Jiacheng Xu, Jessy Li, Greg Durrett.
ACL Findings 2022

Annotating and Modeling Fine-grained Factuality in Summarization
Tanya Goyal, Greg Durrett.
NAACL 2021.

Evaluating Factuality in Generation with Dependency-level Entailment
Tanya Goyal, Greg Durrett.
EMNLP Findings 2020.

Neural Syntactic Preordering for Controlled Paraphrase Generation
Tanya Goyal, Greg Durrett.
ACL 2020.

(short) **Embedding time expressions for deep temporal ordering models**
Tanya Goyal, Greg Durrett.
ACL 2019.

Your behavior signals your reliability: Modeling crowd behavioral traces to ensure quality relevance annotations
Tanya Goyal, Tyler McDonnell, Mucahid Kutlu, Tamer Elsayed, Matthew Lease.
HCOMP 2018.

(short) **An Empirical Analysis of Edit Importance between Document Versions**
Tanya Goyal, Sachin Kelkar, Manas Agarwal, Jeenu Grover.
EMNLP 2017

Preprints

News Summarization and Evaluation in the Era of GPT-3
Tanya Goyal, Jessy Li, Greg Durrett.
arxiv 2022

Shortcomings of Question Answering Based Factuality Frameworks for Error Localization
Ryo Kamoi, **Tanya Goyal**, Greg Durrett.
arxiv 2022 (In submission to EACL)

Understanding Factual Errors in Summarization: Errors, Summarizers, Datasets, Error Detectors
Liyan Tang, **Tanya Goyal**, Alexander Fabbri, Philippe Laban, Jiacheng Xu, Semih Yavuz, Wojciech Kryściński, Justin F. Rousseau, Greg Durrett.
arxiv 2022 (In submission to ARR)

Research Experience

TAUR Lab, UT Austin Fall 2018 - Present
PI: Dr. Greg Durrett

IR and Crowdsourcing Lab, UT Austin Fall 2017 – Spring 2018
PI: Dr. Matt Lease, worked on annotation quality control in crowdsourcing.

Work Experience

Salesforce Research Summer 2021
Host: Wojciech Kryściński, worked on text summarization.

A9 Labs, Amazon Summer 2018
Research Intern
Project: Predictive modelling for customer engagement

Adobe Research, India May 2015 - July 2017
Research Staff
Prototyped predictive models for e-commerce objectives for Adobe Document Cloud

Service

Reviewer (*outstanding reviewer)
EMNLP (2020*, 2021, 2022), CoNLL (2020), NAACL (2021, 2022), ACL (2021, 2022), AKBC (2021), ARR (multiple cycles, 2021-present)

Teaching Assistant, UT Austin Fall 2020
CS378: Natural Language Processing (Undergrad class)

Mentor for Graduate Application Assistance Program, UT Austin 2021-2022
Program to provide assistance during graduate school applications to underrepresented groups

Mentorship

Undergraduate Reading Group, UT Austin Fall 2021, Spring 2022
Lead a small-sized reading group with ~5 undergraduate students covering basics of NLP.

Research Mentorship
- Ryo Kamoi (MS at UT Austin, next Amazon UK) Spring 2022 - Present
- Liyan Tang (PhD at UT Austin) Spring 2022
- Sachin Kelkar (Research Intern at Adobe Research, next MS at Columbia) Summer 2016
- Manas Agarwal (Research Intern at Adobe Research, next MS at Columbia) Summer 2016
- Jeenu Grover (Research Intern at Adobe Research, next Google India) Summer 2016

Invited Talks

NLP Seminar Series at AI Sweden (link) December 2022 (scheduled)
Title: Text summarization and evaluation in the era of GPT-3

Adobe Research India November 2022
Title: Bridging the gap between generation and evaluation capabilities of large language models

University of Alberta July 2019
Title: Syntax-controlled generation using deep neural models

Other Publications

(preprint) **NL-Augmenter: A framework for task-sensitive natural language augmentation**
Kaustubh D Dhole, et al.
arxiv 2021

Contemporary NLP Modeling in Six Comprehensive Programming Assignments
Greg Durrett, Jifan Chen, Shrey Desai, **Tanya Goyal**, Lucas Kabela, Yasumasa Onoe, Jiacheng Xu.
Workshop on Teaching NLP 2021

Preventing inadvertent information disclosures via automatic security policies
Tanya Goyal, Sanket Mehta, Balaji Vasan Srinivasan.
PAKDD 2017

Environment Specific Content Rendering and Transformation
Balaji Vasan Srinivasan, **Tanya Goyal**, Varun Syal, Shubhankar Singh, Vineet Sharma.
IUI 2016 Companion

Patents *(Filed while at Adobe Research, Granted by US PTO)*

Generating data-driven geo-fences	*US Patent 9,838,843*
Method to expand seed keywords into a relevant social query	*US Patent 9,965,766*
Generating predictive models for authoring short messages	*US Patent 10,528,652*
Content to layout template mapping and transformation	*US Patent 10,521,494*
Classifying and ranking changes between document revisions	*US Patent 10,713,432*
Constructing enterprise-specific knowledge graphs	*US Patent 10,915,577*
Determination of Paywall Metrics	*US Patent 10,853,887*
Prediction of tone of interpersonal text communications	*US Patent 10,796,095*
Tagging documents with security policies	*US Patent 10,783,262*
Content optimization for audiences	*US Patent 10,922,492*
Bundling online content fragments for presentation based on content-specific metrics and inter-content constraints	*US Patent 10,891,667*
Method and apparatus for generating predictive insights for authoring short messages	*US Patent 10,073,822*