Cyprus University of Technology
Department of Communication and Internet Studies

**Doctoral Dissertation**

# Instagram hashtags as a source of semantic information for Automatic Image Annotation

**Stamatios Giannoulakis**

**Advisor: Nicolas Tsapatsoulis**

**October 2021**

CYPRUS UNIVERSITY OF TECHNOLOGY

FACULTY OF COMMUNICATION AND MEDIA STUDIES

DEPARTMENT OF COMMUNICATION AND INTERNET STUDIES

Doctoral Dissertation

# Instagram hashtags as a source of semantic information for Automatic Image Annotation

by

Stamatios Giannoulakis

Advisor: Nicolas Tsapatsoulis

Limassol, October 2021

**Approval Form**

Doctoral Dissertation Proposal

# Instagram hashtags as a source of semantic information for Automatic Image Annotation

Presented by

Stamatios Giannoulakis

*Supervisor*: Nicolas Tsapatsoulis, Professor,
Faculty of Communication and Media Studies, Cyprus University of Technology

Signature:

*Committee Member*: Nikos Grammalidis, Researcher B',
Center for Research & Technology Hellas

Signature:

*Committee Member*:Klimis Ntalianis, Professor,
Department of Business Administration, University of West Attica

Signature:

Cyprus University of Technology
Limassol, October 2021

# Copyrights

# Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisor Dr. Nicolas Tsapatsoulis. My supervisor's precious help and support were vital to fulfill my target and finish my Ph.D. thesis—also, the members of the advisory committee, Dr. Lambros Lambrinos, and Dr. Christos Papatheodorou. A special thank you to Eftychios Protopadakis for the valuable help and support. In addition, I wish to thank Christos Rodosthenous for the support and discussions we had. Moreover, I would like to thank Katerina Xenou, who urged me to start the Ph.D., and her support. A special thank you to Vasilis Manavopoulos, and Christos Panayiotou for the journey in the Ph.D.; thank you very much for your support and help all these years. I wish to thank Marilena Agathokleous, Savvas Zenonos, Zenonas Theodosiou, Argyris Argyrou, and all the other colleagues in the Ph.D., for their help and support. I want to thank Dr. Photis Psomopoulos for the help with the latex. Furthermore, I would like to thank my colleagues in the Library and Information Service of the Cyprus University of Technology for their help and support, especially the director Mr. Marios Zervas. I must thank everyone who took the time to fill out the questionnaires for different stages of the research. I also want to thank the examining committee members, Dr. Nikos Grammalidis and Dr. Klimis Ntalianis, for their time and precious comments. Of course, I wouldn't manage to succeed without the valuable assistance of my mother.

The thesis is dedicated to my father, who left us a few years ago.

# Abstract

Billion digital images are uploaded every single day on the Internet and especially on social media. It is vital to develop effective and efficient methods that allow the retrieval of those images according to users' demands. Among the approaches that have been proposed for digital image retrieval is Automatic Image Annotation (AIA). AIA techniques automatically learn the visual representation of semantic concepts from a number of image samples, and use these concept models for tagging new images.

Learning good concept models requires representative pairs of image-tags. Manual annotation is a hard and time-consuming task since a large number of images are necessary to create effective concept models. Moreover, human judgment may contain errors and subjectivity. Therefore, it is highly desirable to find ways for automatically creating training examples, i.e., pairs of images and tags. Contemporary social media, such as Instagram, contain images and associated hashtags, providing a source of indirect annotation. Instagram is a photo-oriented social media platform where users upload images and describe them with hashtags; thus, it might be a rich source for automatically creating pairs of image-tags for AIA.

The thesis focuses on investigating Instagram images and hashtags as a field for AIA purposes. This primary research question is further analyzed through several studies: we define the portion of Instagram hashtags that are related to the visual content of images they accompany and we develop a methodology to locate stophashtags, i.e., common non-descriptive hashtags. We also employ the HITS algorithm in a crowdsourcing environment in order to filter Instagram hashtags and locate the ones that correspond to the visual content of Instagram images they accompany. Topic modelling of Instagram hashtags is introduced as a means for retrieving Instagram images in the traditional text-based information retrieval approach while transfer learning, utilizing filtered Instagram data (pairs of images and hashtags) is applied for a content-based image retrieval scenario.

**Keywords**: Machine Learning, Deep Learning, Automatic Image Annotation, Crowdtagging, Crowdsourcing, Instagram, Hashtags, HITS algorithm, Topic Modelling, Transfer Learning