

THÈSE DE DOCTORAT DE

NANTES UNIVERSITÉ

ÉCOLE DOCTORALE N° 641
*Mathématiques et Sciences et Technologies
de l'Information et de la Communication*
Spécialité : *Informatique*

Par

Manoé KIEFFER

Leveraging annotations in knowledge graphs to enhance their expressiveness for data processing and analysis

Thèse présentée et soutenue à « Lieu », le « date »

Unité de recherche : **Laboratoire des Sciences du Numérique de Nantes**

Thèse N° : « si pertinent »

Composition du Jury :

Attention, en cas d'absence d'un des membres du Jury le jour de la soutenance, la composition du jury doit être revue pour s'assurer qu'elle est conforme et devra être répercutée sur la couverture de thèse

Président :		(à préciser après la soutenance)
Rapporteurs :	M Zoltan MIKLOS	Professeur, Université de Rennes
	M Pierre-Antoine CHAMPIN	Maître de conférences HDR, Université Claude Bernard Lyon 1
Examineurs :	M ^{me} Mélanie DULONG DE ROSNAY	Directrice de recherche CNRS, Centre Internet et Société
	M Philippe LERAY	Professeur, Nantes Université
Dir. de thèse :	M ^{me} Patricia SERRANO ALVARADO	Maîtresse de conférences HDR, Nantes Université
Co-dir. de thèse :	M ^{me} Margo BERNELIN	Chargée de recherche CNRS, Nantes Université

Titre: Leveraging annotations in knowledge graphs to enhance their expressiveness for data processing and analysis

Keywords: Semantic Web, Knowledge graphs, RDF annotations, expressiveness, license, license composition, link prediction, knowledge graph completion

Abstract: Semantic Web standards facilitate the circulation of information online. In particular, RDF knowledge graphs faithfully represent statements and their relations thanks to their great expressiveness, which is further enhanced through the use of RDF annotations. This thesis studies the benefits of annotations in the context of three problems: (1) how to faithfully represent and compose licenses and the intellectual property rights governing online resources; (2) what is the impact of annotations on storing and querying knowledge graphs, and what is the best way to implement

annotations; and (3) how to predict missing information in annotated knowledge graphs. For the first problem, we propose **CLiC**, an explainable algorithm that composes complex licenses into one compliant license. For the second problem, we propose a **benchmarking approach** that determines the best way to implement annotations in a given context. For the third problem, we propose **LP4P**, the first link prediction model that fully supports annotations, and **WD4P**, a benchmark dataset to evaluate such link prediction models.

Titre : Mettre à profit les annotations dans les graphes de connaissances pour améliorer leur expressivité en vue de leur traitement et de leur analyse

Mot clés : Web sémantique, graphe de données, annotations RDF, expressivité, licence, composition de licence, prédiction de liens, complétion de graphe de données

Résumé : Les standards du Web sémantique facilitent la circulation d'information en ligne. En particulier, les graphes de connaissances RDF représentent fidèlement l'information grâce à une grande expressivité, elle-même encore renforcée par les annotations RDF. Cette thèse étudie l'avantage des annotations dans le contexte de trois problèmes : (1) comment représenter et composer des licences et les droits de propriété intellectuelle régissant les ressources en ligne ; (2) quel est l'impact des annotations sur le stockage et l'interrogation des graphes de connaissances, et quelle est la meilleure façon de mettre en œuvre les annotations ; et (3) comment pré-

dire les informations manquantes dans des graphes de connaissances annotés. Pour le premier problème, nous proposons **CLiC**, un algorithme explicable qui permet de composer des licences complexes en une seule licence conforme. Pour le deuxième problème, nous proposons une approche afin de déterminer la meilleure façon de mettre en œuvre les annotations dans un contexte donné. Pour le troisième problème, nous proposons **LP4P**, le premier modèle de prédiction de liens qui prend entièrement en charge les annotations, et **WD4P**, un ensemble de données de référence permettant d'évaluer ces modèles de prédiction de liens.