

UTILISER LES MAJUSCULES D'IMPRIMERIE

A REMPLIR PAR L'ETUDIANT

UNIVERSITE ou
ETABLISSEMENT EURECOM -TELECOM PARISTECH
AUTEUR
NOM DE NAISSANCE ATEMEZING
PRENOM(S) AUGUSTE GHISLAIN
NOM D'USAGE ATEMEZING
NOM FIGURANT SUR LA PUBLICATION, si différent des deux précédents
DATE DE NAISSANCE (*) (Format JJMMAAAA) 04051979

ADRESSE(*)
N° et RUE 29, RUE CASTERES
CODE POSTAL 92110 Pays FRANCE
VILLE CLICHY
ADRESSE ELECTRONIQUE(*) (facultatif) GHISLAIN.ATEMEZING@GMAIL.COM

DATE DE SOUTENANCE ANNEE (AAAA) 2015 MOIS (MM) 04

DIRECTEUR DE THESE
SPECIALITE INFORMATIQUE
NOM TRONCY
PRENOM (S) RAPHAEL

TYPE DE DOCTORAT : COCHER LE TYPE DE DOCTORAT

☒ THESE DOCTORAT ☐ THESE D'EXERCICE ☐ THESE DOCTORAT D'ETAT

DISCIPLINE Science de l'information de la communication

INDICATIONS BIBLIOGRAPHIQUES :

Nombre de volumes : 1 Nombre total de pages : 261 Nombres de références bibliographiques : 212

AUCUN TEXTE MANUSCRIT N'EST ADMIS. DACTYLOGRAPHIER OU COLLER LE TEXTE IMPRIME DANS LES ZONES PREVUES.

TITRE DE LA THESE EN FRANÇAIS : Transcrire en toutes lettres les symboles spéciaux.

Interactions publier et consommer des données gouvernementales sur le web sémantique
--

(*) Les informations demandées sur ce formulaire sont destinées, à l'exception de l'adresse, à la constitution du catalogue collectif du Sudoc. Le Sudoc a fait l'objet d'une déclaration à la CNIL (arrêté du 10 septembre 1991 : publié au Journal Officiel du 1^{er} octobre 1991). La déclaration est obligatoire en application de l'arrêté du 25 septembre 1985 relatif au dépôt, signalement et reproduction des thèses ou travaux présentés en soutenance en vue du doctorat. Le droit d'accès et de rectification prévu par la loi n° 78 – 17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés ne concerne que les informations de la première page de ce formulaire. Il peut s'exercer auprès de : Agence bibliographique de l'enseignement supérieur – BP 4367 / 34196 MONTPELLIER CEDEX 5.

RESUME DE LA THESE EN FRANCAIS

Chaque cadre doit contenir un résumé de 1700 caractères maximum, espaces compris. En cas de dépassement, la coupure sera automatique.

Le doctorant adressera son texte sous forme électronique (disquette, etc.)

Pour les modalités pratiques, contactez votre bibliothèque.

Dans cette thèse, nous explorons comment l'utilisation des standards et des technologies du web sémantique peut aider à résoudre les problèmes causés par l'hétérogénéité et la diversité des formats de données et des structures de représentation dans le domaine géographique.

Cette thèse applique les principes des «données liées» dans le domaine de l'information géographique. Nous traitons 3 aspects essentiels: (1) Comment représenter efficacement et stocker des données géospatiales sur le Web? (2) Comment interagir avec des données sémantiques interconnectées ? (3) Quels mécanismes peuvent aider la préservation des données structurées de haute qualité sur le Web?

Nous proposons et développons 3 vocabulaires pour représenter des systèmes de coordonnées de référence (CRS), des entités topographiques et la géométrie associée à ces entités. Ces ontologies étendent d'autres vocabulaires existants et ajoutent deux avantages supplémentaires : l'utilisation explicite de CRS identifiés par des URIs pour représenter la géométrie, et la capacité de décrire des géométries structurées en RDF. Nous avons ainsi publié la base de données GEOFLA.

Concernant la «consommation» de données RDF, nous proposons un vocabulaire pour décrire les applications de visualisation (DVIA). En outre, nous formalisons et mettons en œuvre un workflow pour visualiser des données sémantiques interconnectées à travers l'outil LDVizWiz, un assistant de visualisation générique de données liées sur le Web.

La dernière partie de la thèse décrit nos contributions au catalogue des vocabulaires liés (LOV) ainsi qu'une proposition originale pour la réutilisation des vocabulaires.

RESUME DE LA THESE EN ANGLAIS

In this thesis, we explore how semantic web technologies can be used to tackle the research problems related to the integration and consumption of geo-spatial data.

This thesis applies the Linked Data principles in the domain of geographic information (a key domain for open government). In particular, we address three key challenging problems in the publishing workflow of geospatial open data publication and consumption, with real world use cases from the the French National mapping agency (IGN): (1) How to efficiently represent and store geospatial data on the Web to ensure interoperable applications? (2) What are the best options for a user to interact with semantic content using visualizations? (3) What are the mechanisms that support preserving structured data of a high quality on the Web?

Our contributions are thus break down into three parts with applications in the geographical domain. We propose and model four vocabularies for representing coordinate reference systems (CRS), topographic entities and their geometries. These ontologies extend existing vocabularies and add two additional advantages: an explicit use of CRS identified by URIs for geometry, and the ability to describe structured geometries in RDF.

Regarding the consumption of linked data, we propose a vocabulary for Describing Visualization Applications (DVIA). We formalize and implement a novel workflow for visualizing datasets with the LDVizWiz tool: a Linked Data Visualization Wizard.

The last part of the thesis describes contributions to the Linked Open Vocabularies (LOV) catalogue: it shows how LOV can be used with an ontology modeling methodology to improve reuse of vocabularies.

TITRE DE LA THESE EN ANGLAIS : Transcrire en toutes lettre les symboles spéciaux

Publishing and Consuming Geo-Spatial and Government Data on the Semantic Web

PROPOSITION DE MOTS-CLES : ces termes décrivant le contenu de la thèse seront saisis dans le Sudoc.

Data on the Web
Web
Semantic Web
Linked Data

vocabularies
geospatial data
Visualization
Semantic publishing

DATE ET SIGNATURE DE L'AUTEUR

--

A REMPLIR PAR LE SERVICE DE DOCTORAT

UNIVERSITE ou
ETABLISSEMENT

AUTORISATION DE REPRODUCTION PAR LE JURY

☐

OUI

☐

NON

☐

NON : thèse non reproduite car non corrigée dans les 3 mois réglementaires

☐

NON : thèse confidentielle **JUSQU'EN**

APRES VERIFICATION DATE et CACHET DU SERVICE DE DOCTORAT

A REMPLIR PAR LA BIBLIOTHEQUE

CATALOGAGE DE LA THESE ASSURE PAR LA BIBLIOTHEQUE DE L'UNIVERSITE OU DE L'ETABLISSEMENT

LOCALISATION DE L'EXEMPLAIRE ORIGINAL

**RBCCN
NUMERO D'IDENTIFICATION**

Composé de 12 caractères : les 4 premiers indiquent l'année de soutenance, les 4 suivants définissent l'établissement, les 4 derniers correspondent à un numéro séquentiel

DISCIPLINE (telle qu'elle est indiquée sur la page de titre de la thèse.)

INDEXATION RAMEAU

1

2

3

4

5

INDEXATION LIBRE

RAPPEL : SI LE JURY A AUTORISE LA REPRODUCTION DE LA THESE, UN EXEMPLAIRE DE LA THESE ET DU FORMULAIRE SERONT TRANSMIS A L'ATELIER NATIONAL DE REPRODUCTION DES THESES (ANRT) DE GRENOBLE OU DE LILLE.

DATE ET CACHET DE L'ETABLISSEMENT