On Combining Alignment Techniques

Tordai, A.

2012

document version
Publisher's PDF, also known as Version of record

Samenvatting

Karakteristiek voor collecties van instellingen voor cultureel erfgoed is dat ze bestaan uit (afbeeldingen van) culturele objecten, hun beschrijving in de vorm van metadata, en bijbehorende woordenlijsten of vocabulaires. Deze laatste worden gebruikt voor het indexeren van metadata, waardoor het zoeken naar objecten in de collectie eenvoudiger wordt. Collecties variëren niet alleen in hun inhoud, maar maken ook gebruik van vele verschillende metadataformaten doordat de metadata onafhankelijk worden samengesteld door elke instelling. Dit bemoeilijkt de integratie van verschillende collecties in een enkele virtuele collectie. De metadata en vocabulaires moeten eerst omgezet worden in een gemeenschappelijk formaat en vervolgens aan elkaar gekoppeld worden.

In dit proefschrift beginnen wij met het onderzoeken van de stappen die nodig zijn voor de integratie van een collectie in een meeromvattende virtuele collectie. We identificeren vier verschillende stappen en illustreren ze met behulp van een case study. Ten eerste zetten we alle vocabulaires die bij de metadata horen om in een gemeenschappelijk formaat (SKOS). Ten tweede zetten we de metadata schemata (de verzameling elementen die de metadata beschrijven) om in generieke schema's (Dublin Core en VRA). Ten derde zetten we de waarden van de metadata om. In deze stap proberen we zinvolle waarden te identificeren die ofwel afkomstig zijn uit vocabulaires ofwel zelf identificeerbare objecten zijn, zoals een schilder wiens werken opgenomen zijn in verschillende collecties. Als beide niet mogelijk zijn, houden we als laatste optie waarden aan als tekst. In de vierde en laatste stap beelden we collectievocabulaires af op vocabulaires die al deel uitmaken van de virtuele collectie; met andere woorden, we maken verbindingen tussen vergelijkbare concepten. Hierdoor wordt de collectie die geïndexeerd is met het afgebeelde vocabulaire geïntegreerd met de andere collecties. De vocabulaires kunnen handmatig (een tijdrovende taak) of automatisch afgebeeld worden op andere vocabulaires. Er zijn vele tools beschikbaar die gebruik maken van één of meerdere technieken om een afbeelding te maken. Hoewel deze tools geëvalueerd en vergeleken worden in de jaarlijkse Ontology Alignment Evaluation Initiative (OAEI) is er geen methodologisch advies over hoe vocabulaires van begin tot eind verbonden kunnen worden.

In het vervolg van het proefschrift richten we ons op vocabulaire-afbeelding. Onze belangrijkste onderzoeksvraag is als volgt: textit Hoe kunnen we afbeeldingstechnieken combineren, hun prestatie beoordelen en de resulterende afbeelding evalueren?

In de hoofdstukken 3 en 4 richten we ons op het eerste deel van de onderzoeksvraag, en bestuderen we het combineren van afbeeldingstechnieken voor vocabulaires en het beoordelen van hun prestaties. In ons experiment in hoofdstuk 3 worden twee vocabulaires afgebeeld met behulp van tekstvergelijking ("string matching") en kant en klare afbeeldingstools. We concluderen dat de

afzonderlijke tools relatief slecht presteren; echter, door het uitkiezen van specifieke deelverzamelingen is de kwaliteit van gegenereerde verbindingen in de deelverzameling hoger dan die in de oorspronkelijke verzameling.

In hoofdstuk 4 voeren we een vergelijkbaar experiment uit, maar dan op grotere vocabulaires. Vanwege de omvang van de vocabulaires waren we niet in staat om de kant en klare tools uit hoofdstuk 3 te gebruiken. Daarom kozen we voor eenvoudige tekstvergelijkingstechnieken om verbindingen te maken. Doordat het aantal gegenereerde verbindingen te groot is konden wij geen handmatige evaluatie toepassen op de gehele verzameling van verbindingen. Om de waarde van elke techniek te kunnen beoordelen namen wij steekproeven van de verbindingen. In een stap-voor-stap methode selecteren wij deelverzamelingen van verbindingen die onderling van hogere kwaliteit zijn dan de oorspronkelijke afbeelding.

In hoofdstuk 5 richten we ons op het hergebruik van verbindingen tussen gelijkwaardige concepten om verbindingen tussen meerdere vocabulaires samen te stellen. Wij testen of gelijkwaardige verbindingen overdraagbaar zijn in de praktijk door experimenten uit te voeren op vocabulaires uit meerdere domeinen: geneeskunde, cultureel erfgoed en het bibliotheekdomein. We genereren samengestelde verbindingen en onderzoeken ze op hun aantal en hun kwaliteit. Onze bevindingen kwamen overeen met onze verwachtingen, met name dat samengestelde verbindingen tussen vocabulaires die hetzelfde domein beschrijven van hogere kwaliteit zijn dan verbindingen tussen vocabulaires uit verschillende domeinen. Een verrassende bevinding die meer onderzoek verdient is dat de kwaliteit van samengestelde verbindingen hoger is dan verwacht, namelijk vergelijkbaar met de kwaliteit van de gebruikte verbindingen.

Voor het beoordelen van de kwaliteit van een afbeelding is het nodig om tenminste een gedeelte handmatig te beoordelen. Om zeker te zijn dat de handmatige evaluatie op zijn beurt betrouwbaar is, dient tenminste één subset van verbindingen door onafhankelijke beoordelaars te worden geëvalueerd. Hierna wordt de interbeoordelaarsbetrouwbaarheid berekend. Als deze voldoende is kunnen de handmatig geëvalueerde verbindingen gebruikt worden als testverzameling of ?reference alignment?. In onze voorgaande experimenten vonden we dat de interbeoordelaarsbetrouwbaarheid vrij laag uitvalt. In hoofdstuk 6 en 7 bestuderen we de mate van overeenstemming tussen meerdere beoordelaars in verschillende experimenten, en we proberen de oorzaken van lage niveaus van overeenkomst te identificeren. Wij concluderen dat betere richtlijnen met een duidelijke beschrijving van de evaluatietaak en de evaluatiecategorieën de mate van overeenstemming tussen beoordelaars aanzienlijk verbeteren. Toch zorgen bepaalde eigenschappen van handmatige evaluatie ervoor dat dit een moeilijke taak blijft, met name de intrinsieke dubbelzinnigheid van begrippen en hun beschrijvingen, en het feit dat de achtergrondkennis van beoordelaars hun interpretatie van begrippen beïnvloedt. Daarnaast hebben kenmerken van vocabulaires, zoals hun representatie en hun domein, en de gekozen evaluatiecategorieën ook invloed op de uitkomst van een evaluatie.

In dit proefschrift laten we zien dat het combineren van afbeeldingstechnieken door middel van een interactief proces een effectieve en transparante methode is voor het genereren van verbindingen van hoge kwaliteit. We tonen ook aan dat een groot aantal factoren handmatige beoordeling

beïnvloedt. Deze factoren zijn onder andere de kenmerken van vocabulaires, de gekozen evaluatiecategorieën en de inherente vaagheid van concepten. Er moet rekening gehouden worden met deze factoren bij het verbinden van vocabulaires.

We sluiten af met een methode voor het evalueren van verbindingen waarbij onder andere de overeenstemmingsniveaus tussen de beoordelaars gerapporteerd worden, samen met een rapport over de manier waarop meningsverschillen werden behandeld. Hierbij wordt het belang van het openbaar maken van de evaluatie en het rapporteren daarover benadrukt. Daarnaast stellen we een methode voor voor het verbinden van vocabulaires. Deze methode maakt het mogelijk om de oorsprong van verbindingen te traceren doordat het koppelen in expliciete stappen gebeurt, waardoor verbindingen van bepaalde kwaliteit geselecteerd kunnen worden. Hierdoor kan het verbindingsproces aangepast worden aan verschillende toepassingen.