



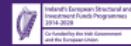
Engaging Content
Engaging People

Linking Historical Sources to Established Knowledge Bases in Order to Inform Entity Linkers in Cultural Heritage

Gary Munnelly Annalina Caputo Séamus Lawless



The ADAPT Centre is funded under the SFI Research Centres Programme(Grant 13/RC/2106)and is co-funded under the European Regional Development Fund.



Overarching research looks at Entity Linking in Irish Cultural Heritage collections.

Many possible advantages to being able to perform reliable Entity Linking on digital cultural heritage collections:

- Helps to organise content.
- Integrates collections with existing resources.
- Enables us to build bridges between disparate collections based on mutual entities.
- Facilitates the construction of better search and/or personalization tools.

⋮



Entities in Cultural Heritage collections often have poor representation in popular Knowledge Bases

Study	Content	Coverage ¹
Agirre et al.	Europeana Scran and Cgrid collections	22%
Munnelly et al.	17 th Century English Legal	23%

An investigation by Max de Wilde reported recall scores between 0.213 and 0.517 for data curated by Historische Kranten.



¹Studies used Wikipedia derived source for Knowledge Base

Find a better knowledge base

1. GeoNames
2. Pleiades
3. Getty
4. Library of Congress



Find a better knowledge base

1. GeoNames
2. Pleiades
3. Getty
4. Library of Congress

We couldn't find a source with adequate coverage for the collections we handle.



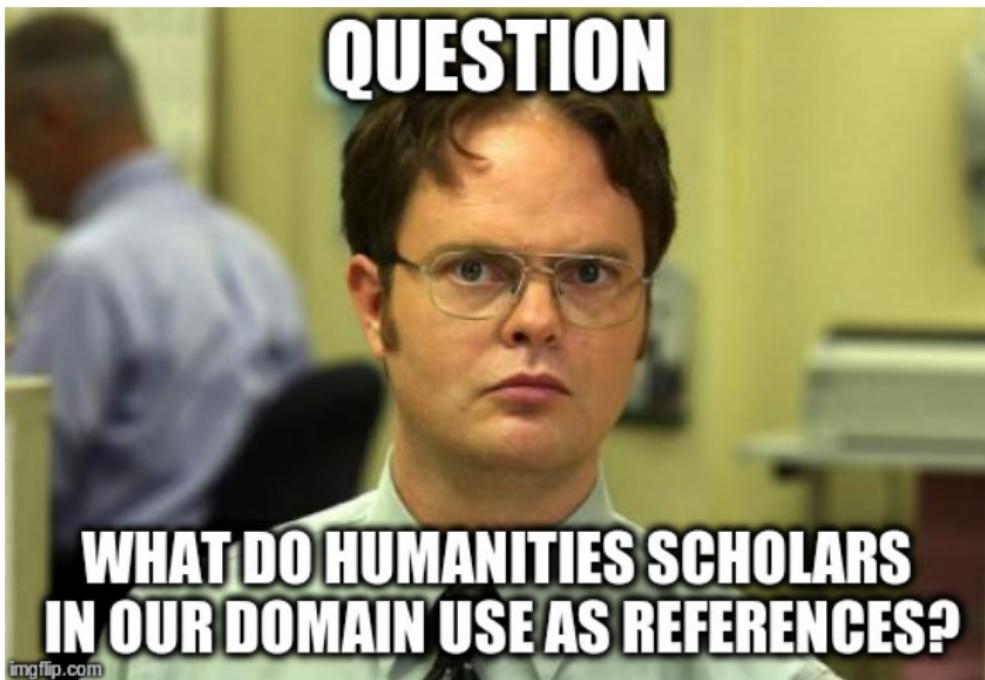
Find a better knowledge base

1. GeoNames
2. Pleiades
3. Getty
4. Library of Congress

We couldn't find a source with adequate coverage for the collections we handle.

Can we build our own knowledge base?





- Statute Staple
- Petty Maps
- Books of Survey and Distribution
- Lodges Peerage



- Statute Staple
- Petty Maps
- Books of Survey and Distribution
- Lodges Peerage

Excellent sources of information, but very isolating. Essentially reduces our efforts to little more than record resolution.
Guaranteed to be noisy and open to dispute.



- Oxford Dictionary of National Biography (ODNB)
- Dictionary of Irish Biography (DIB)



- Oxford Dictionary of National Biography (ODNB)
- Dictionary of Irish Biography (DIB)

Not bad! Coverage is not as good as the primary sources, but entities are more concrete.

Still somewhat isolating. Can we connect these resources to a more established knowledge base?



Identify corresponding DBpedia entities for DIB and ODNB bibliographies:

- Enables us to link our collections with more accepted linked data resources.
- Implicitly limits the Entity Linker to entities that are relevant to our geographic region.
- Enables linking across multiple knowledge bases – shown to be beneficial by Brando et al.



Given a bibliography which describes an entity:

1. Retrieve DBpedia candidates which may discuss same entity.
2. Compare bibliography to all candidates.
3. Choose most similar candidate, or disregard all if no suitable equivalent is found.
4. Link bibliography to corresponding DBpedia entry.



Method: Features

www.adaptcentre.ie

Coote, Sir Charles

by Robert Armstrong

Coote, Sir Charles (d. 1642), army officer, was the eldest son of Sir Nicholas Coote of Blownorton, Norfolk, and had two sisters and at least two brothers, one of whom, William, became dean of Down. His mother was almost certainly Nicholas Coote's second wife, Eleanor, daughter of Sir Michael Stanhope and his wife, Anne, and sister of the courtier Sir John Stanhope, who sponsored Charles Coote's early career in Ireland, perhaps from 1600. Coote held command of an infantry company in Munster from 1601 until some time after 1603. He was granted a reversion to the post of provost marshal of Connacht in June 1605, apparently took up that office soon afterwards, and was appointed one of the collectors of composition rents for Connacht in November 1613. In 1615 a royal order for the confirmation of land titles in the province prompted an investigative commission, on which Coote served, and the payment of large sums of money by property holders, but the failure to fully follow through with the enrolment of surrenders and grants, often blamed on clerical negligence, was also attributed to Coote's 'calculated remissness' (Clarke, 53). Certainly he proceeded to build up extensive landed possessions in the province, in counties Leitrim, Roscommon and Sligo. By 1617 he had married Dorothea, probably the younger daughter of Hugh Cuffe, plantation undertaker in Co. Cork, who brought property in Co. Cork and in Queen's County to the marriage. They had four sons and one daughter; the eldest son, also Charles (qv), became 1st earl of Mountrath.

Coote, Sir Charles, first baronet

(d. 1642)

Pádraig Lenihan

<https://doi.org/10.1093/ref:odnb/6239> Published in print: 23 September 2004

Published online: 23 September 2004

Coote, Sir Charles, first baronet (d. 1642), soldier and political administrator, was the elder of the two sons of Sir Nicholas Coote and his second wife, Eleanor Stanhope. Sir Nicholas had been a retainer of the duke of Norfolk, in whose cause he struck the attorney-general, for which he was fined £20,000. He spent some time in prison and at last was obliged (with his son's permission) to sell off part of his estates to pay the fine.

Charles is said to have been at Cambridge in 1591, and then served as a captain under Charles Blount, Lord Mountjoy, in 1600–03, the concluding three years of the Nine Years' War; in 1605 he was appointed provost-marshal of Connacht for life. His marriage about 1610 to Dorothea, daughter of Hugh Cuffe of Cuffeswood, co. Cork, brought them four sons and a daughter.

Sir Charles Coote, 1st Baronet

From Wikipedia, the free encyclopedia

Sir Charles Coote, 1st Baronet (1581 – May 1642) was an English soldier, administrator and landowner who lived in [Ireland](#).

He was born into a Devonshire family, the son of Sir Nicholas Coote. In 1600 he moved to Ireland as a captain of the 100th Foot Regiment in the army of Lord Mountjoy, Lord Deputy of Ireland, where he fought in the last few years of the [Nine Years War](#) and was at the Siege of Kinsale in 1601–02 which ultimately led to the defeat of the O'Neills.

In 1605 he was appointed Provost-Marshal of Connaught for life and in 1613 was appointed to the office of General Collector and Receiver of the King's Composition Money for Connaught, also for life. In 1620 he was promoted to vice-President of Connaught and made a baronet (of Castle Cuffe in Queen's County) the following year. He became a substantial landowner and served as a commissioner to examine and contest Irish land titles. In 1634 he was appointed *Custos Rotulorum* of Queen's County, again for life. He was elected Member of Parliament (M.P.) in the [Parliament of Ireland](#) for Queen's County in 1639.



circle of Cornelius Johnson, Sir Charles Coote, 1st Baronet, ca. 1630



Method: Features

www.adaptcentre.ie

Coote, Sir Charles

by Robert Armstrong

Coote, Sir Charles (d. 1642), army officer, was the eldest son of Sir Nicholas Coote of Blownorton, Norfolk, and had two sisters and at least two brothers, one of whom, William, became dean of Down. His mother was almost certainly Nicholas Coote's second wife, Eleanor, daughter of Sir Michael Stanhope and his wife, Anne, and sister of the courtier Sir John Stanhope, who sponsored Charles Coote's early career in Ireland, perhaps from 1600. Coote held command of an infantry company in Munster from 1601 until some time after 1603. He was granted a reversion to the post of provost marshal of Connacht in June 1605, apparently took up that office soon afterwards, and was appointed one of the collectors of composition rents for Connacht in November 1613. In 1615 a royal order for the confirmation of land titles in the province prompted an investigative commission, on which Coote served, and the payment of large sums of money by property holders, but the failure to fully follow through with the enrolment of surrenders and grants, often blamed on clerical negligence, was also attributed to Coote's 'calculated remissness' (Clarke, 53). Certainly he proceeded to build up extensive landed possessions in the province, in counties Leitrim, Roscommon and Sligo. By 1617 he had married Dorothea, probably the younger daughter of Hugh Cuffe, plantation undertaker in Co. Cork, who brought property in Co. Cork and in Queen's County to the marriage. They had four sons and one daughter; the eldest son, also Charles (qv), became 1st earl of Mountrath.

Coote, Sir Charles, first baronet

(d. 1642)

Pádraig Lenihan

<https://doi.org/10.1093/ref:odnb/6239> Published in print: 23 September 2004

Published online: 23 September 2004

Coote, Sir Charles, first baronet (d. 1642), soldier and political administrator, was the elder of the two sons of Sir Nicholas Coote and his second wife, Eleanor Stanhope. Sir Nicholas had been a retainer of the duke of Norfolk, in whose cause he struck the attorney-general, for which he was fined £20,000. He spent some time in prison and at last was obliged (with his son's permission) to sell off part of his estates to pay the fine.

Charles is said to have been at Cambridge in 1591, and then served as a captain under Charles Blount, Lord Mountjoy, in 1600–03, the concluding three years of the Nine Years' War; in 1605 he was appointed provost-marshal of Connacht for life. His marriage about 1610 to Dorothea, daughter of Hugh Cuffe of Cuffeswood, co. Cork, brought them four sons and a daughter.

Sir Charles Coote, 1st Baronet

From Wikipedia, the free encyclopedia

Sir Charles Coote, 1st Baronet (1581 – May 1642) was an English soldier, administrator and landowner who lived in Ireland.

He was born into a Devonshire family, the son of Sir Nicholas Coote. In 1600 he moved to Ireland as a captain of the 100th Foot Regiment in the army of Lord Mountjoy, Lord Deputy of Ireland, where he fought in the last few years of the Nine Years War and was at the Siege of Kinsale in 1601–02 which ultimately led to the defeat of the O'Neills.

In 1605 he was appointed Provost-Marshal of Connaught for life and in 1613 was appointed to the office of General Collector and Receiver of the King's Composition Money for Connaught, also for life. In 1620 he was promoted to vice-President of Connaught and made a baronet (of Castle Cuffe in Queen's County) the following year. He became a substantial landowner and served as a commissioner to examine and contest Irish land titles. In 1634 he was appointed *Custos Rotulorum* of Queen's County, again for life. He was elected Member of Parliament (M.P.) in the Parliament of Ireland for Queen's County in 1639.



circle of Cornelius Johnson, Sir Charles Coote, 1st Baronet, ca. 1630



Method: Features

www.adaptcentre.ie

Coote, Sir Charles

by Robert Armstrong

Coote, Sir Charles (d. 1642), army officer, was the eldest son of Sir Nicholas Coote of Blownorton, Norfolk, and had two sisters and at least two brothers, one of whom, William, became dean of Down. His mother was almost certainly Nicholas Coote's second wife, Eleanor, daughter of Sir Michael Stanhope and his wife, Anne, and sister of the courtier Sir John Stanhope, who sponsored Charles Coote's early career in Ireland, perhaps from 1600. Coote held command of an infantry company in Munster from 1601 until some time after 1603. He was granted a reversion to the post of provost marshal of Connacht in June 1605, apparently took up that office soon afterwards, and was appointed one of the collectors of composition rents for Connacht in November 1613. In 1615 a royal order for the confirmation of land titles in the province prompted an investigative commission, on which Coote served, and the payment of large sums of money by property holders, but the failure to fully follow through with the enrolment of surrenders and grants, often blamed on clerical negligence, was also attributed to Coote's 'calculated remissness' (Clarke, 53). Certainly he proceeded to build up extensive landed possessions in the province, in counties Leitrim, Roscommon and Sligo. By 1617 he had married Dorothy, probably the younger daughter of Hugh Cuffe, plantation undertaker in Co. Cork, who brought property in Co. Cork and in Queen's County to the marriage. They had four sons and one daughter; the eldest son, also Charles (qv), became 1st earl of Mountrath.

Coote, Sir Charles, first baronet

(d. 1642)

Pádraig Lenihan

<https://doi.org/10.1093/ref:odnb/6239> Published in print: 23 September 2004

Published online: 23 September 2004

Coote, Sir Charles, first baronet (d. 1642), soldier and political administrator, was the elder of the two sons of Sir Nicholas Coote and his second wife, Eleanor Stanhope. Sir Nicholas had been a retainer of the duke of Norfolk, in whose cause he struck the attorney-general, for which he was fined £20,000. He spent some time in prison and at last was obliged (with his son's permission) to sell off part of his estates to pay the fine.

Charles is said to have been at Cambridge in 1591, and then served as a captain under Charles Blount, Lord Mountjoy, in 1600–03, the concluding three years of the Nine Years' War; in 1605 he was appointed provost-marshal of Connacht for life. His marriage about 1610 to Dorothy, daughter of Hugh Cuffe of Cuffeswood, Co. Cork, brought them four sons and a daughter.

Sir Charles Coote, 1st Baronet

From Wikipedia, the free encyclopedia

Sir Charles Coote, 1st Baronet (1581 – May 1642) was an English soldier, administrator and landowner who lived in [Ireland](#).

He was born into a Devonshire family, the son of Sir Nicholas Coote. In 1600 he moved to Ireland as a captain of the 100th Foot Regiment in the army of Lord Mountjoy, Lord Deputy of Ireland, where he fought in the last few years of the [Nine Years War](#) and was at the Siege of Kinsale in 1601–02 which ultimately led to the defeat of the O'Neills.

In 1605 he was appointed Provost-Marshal of Connaught for life and in 1613 was appointed to the office of General Collector and Receiver of the King's Composition Money for Connaught, also for life. In 1620 he was promoted to vice-President of Connaught and made a baronet (of Castle Cuffe in Queen's County) the following year. He became a substantial landowner and served as a commissioner to examine and contest Irish land titles. In 1634 he was appointed *Custos Rotulorum* of Queen's County, again for life. He was elected Member of Parliament (M.P.) in the [Parliament of Ireland](#) for Queen's County in 1639.



circle of Cornelius Johnson, Sir Charles Coote, 1st Baronet, ca. 1630



- Indexed all DBpedia people in Solr instance.
- Indexed content included.
 - ▶ Name of DBpedia entity.
 - ▶ Text of corresponding Wikipedia article.
 - ▶ Anchor text associated with entity.
- For candidate retrieval, title of biography was executed as query against search index.
 - ▶ Results boosted on matches in title and anchor text.
- Method only looks at top 10 returned results.



Compared title of biography to name of DBpedia entity using Monge-Elkan distance:

- Tokenize both strings and add to bipartite graph.
- Find optimal mapping between sets of tokens based on string similarity:
 - ▶ Similarity computed with Jaro-Winkler distance.
- Compute harmonic mean of similarities based on optimal mapping.



Can be vocabulary differences between biographies and Wikipedia.

Use Word Mover Similarity (WMS) to compare content of articles:

- Trained a Word2Vec model on contents of Wikipedia.
- Convert content of biography and content of all candidate entity biographies to vector representation using Word2Vec Model.
- Compute similarity between articles by computing how far apart they are in space.



Final score is a linear combination of the two previous scores:

$$\Psi(b, p) = \alpha \Phi(b_{title}, p_{name}) + \beta \Omega(b_{content}, p_{article})$$

Where α and β are tuning parameters.

We apply a hard threshold to this score. If the similarity of the highest ranked candidate is less than this score, then the bibliography does not have a corresponding DBpedia entry.

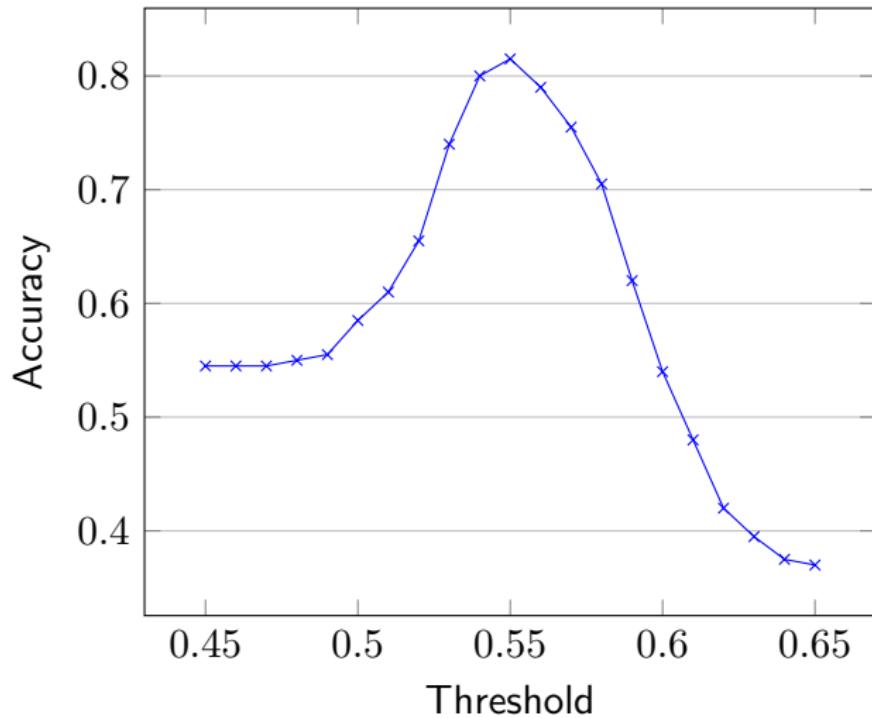


- Sampled 200 bibliographies from ODNB and DIB:
 - ▶ 400 samples in total.
- Created manually annotated gold standard from DBpedia:
 - ▶ 72 DIB bibliographies do not have DBpedia referent.
 - ▶ 64 ODNB bibliographies do not have DBpedia referent.
- Presented approach is effectively an Entity Linking method:
 - ▶ Evaluated with BAT Framework.
- Results given in terms of percentage correct vs. incorrect annotations on the gold standard.



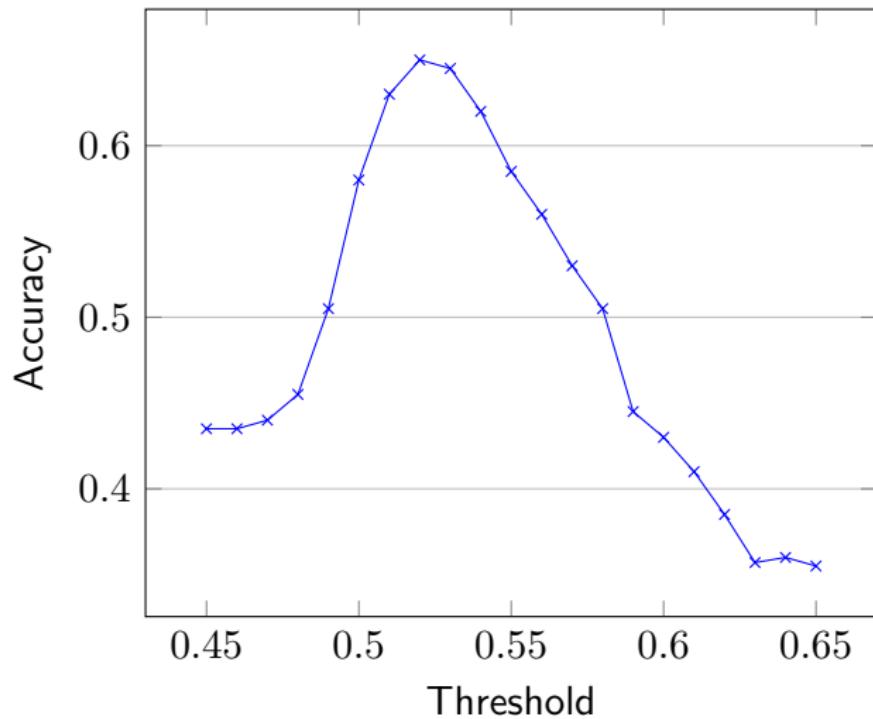
Results: Threshold DIB

www.adaptcentre.ie



Results: Threshold ODNB

www.adaptcentre.ie



- Wide disparity in performance on two collections:
 - ▶ 81.5% accurate on DIB.
 - ▶ 67.5% accurate on ODNB.
- Inaccuracies due to Solr not finding referent as candidate:
 - ▶ 45.9% of errors on DIB.
 - ▶ 43.1% of errors on ODNB.
- Problems with ODNB likely due to pictorial resources:
 - ▶ very little text content for WMS to compare.



- Presented a method for linking bibliographies to DBpedia.
- Resulting output can be used as knowledge base for Entity Linking.
- Links to DBpedia enable us to associate our historical collections with established knowledge bases where possible.
- Want to investigate the possible benefits of linking across multiple knowledge bases, given that no single resource satisfies our requirements.



Thank You Questions



Eneko Agirre et al. "Matching Cultural Heritage items to Wikipedia". In: *LREC*. 2012, pp. 1729–1735.

Carmen Brando, Francesca Frontini, and Jean-Gabriel Ganascia. "REDEN: Named Entity Linking in Digital Literary Editions Using Linked Data Sets". In: *Complex Systems Informatics and Modeling Quarterly* 7 (July 2016), pp. 60–80.

Marco Cornolti, Paolo Ferragina, and Massimiliano Ciaramita. "A framework for benchmarking entity-annotation systems". In: *Proceedings of the 22nd international conference on World Wide Web*. ACM. 2013, pp. 249–260.

Sergio Jimenez et al. "Generalized mongue-elkan method for approximate text string comparison". In: *International Conference on Intelligent Text Processing and Computational Linguistics*. Springer. 2009, pp. 559–570.



Matt Kusner et al. "From word embeddings to document distances". In: *International Conference on Machine Learning*. 2015, pp. 957–966.

Alvaro Monge and Charles Elkan. "The field matching problem: Algorithms and applications". In: *In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. 1996, pp. 267–270.

Gary Munnelly and Séamus Lawless. "Investigating Entity Linking in Early English Legal Documents". In: *ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. In Press.

