Mapping-Chains for studying Concept Shift in Political Ontologies

Shenghui Wang[1 2], Stefan Schlobach[2], Janet Takens[1], and Wouter van Atteveldt[1]

[1] Department of Communication Science
[2] Department of Computer Science
Vrije Universiteit Amsterdam

Abstract. For some years now ontologies have been used in Social Science, *e.g.* , in annotation of newspaper articles for disambiguating concepts within Media Analysis. These ontologies and annotations have now become objects of study in their own right, as they implicitly represent the shift of meaning of political concept over time. Manual mappings, which are intrinsically intensional, can hardly capture such subtle changes, but we claim that automatic instance-based mappings, with their extensional character, are more suitable for producing interesting **mapping-chains.**
In this paper, we evaluate the use of instance-based ontology mappings for producing concept chains in a case-study in Communication Science on a corpus with ontologies describing the Dutch election campaigns since 1994. This initial research shows the potential of the associative character of extensional mapping-chains, but also indicates a number of unsolved open questions, most significantly the lack of a proper methodology for evaluating such chains due to the open, explorative character of the task.

1 Introduction

Since 1994 Communications Scientists at the Vrije Universiteit Amsterdam have been annotating newspaper articles with controlled vocabularies (of increasing expressiveness) quantitatively studying the influence of the Media on the political processes. The idea is to code the *meaning* of sentences and articles in a formalised graph representation (called NET) similar to RDF triples. In these triples actors and issues are taken from an ontology, and the predicate usually represents opinions and moods. During recent election campaigns all newspaper articles on Dutch politics were manually coded using the NET method, with a different ontology used in each of the elections. Each ontology is more or less an adaptation of a previous one, with different foci, as in each election new issues emerged and the (societies' and scientists') view on issues changed.

As now several of these campaign data sets can easily be queried, also through the use of Semantic Web technology, Communication Scientists' interests started to also include temporal shifts of political development. In an initial analysis political developments over time were studied by querying the NET representation

of the articles from the different campaigns, which required manual mappings between the ontologies. In this paper, we propose a different approach, namely to study *concept shift* by using chains of extensional, *i.e.* , instance-based, mappings. Our hypothesis is that these mapping-chains represent subtle changes in meaning of the related concepts over time, and in this paper we will investigate this claim.

Methodology Following our previous work [1] we use information retrieval techniques to calculate document similarity between annotated articles, which we use subsequently to identify similarity between concepts. Chains between the most similar of these concepts thus produce graph-like structures (lines, trees, or DAGs), which "tell their own" story of Dutch politics over the past 15 years.

Research questions There are two major research questions, one regarding the correctness of our hypothesis, the second concerning the validity of our approach. More concretely we have to investigate:

- **RQ1:** What are suitable structures for representing mapping chains?
- **RQ2:** Can instance-based ontology mapping provide *useful* mapping-chains expressing concept shift, and how can we evaluate those chains?

The second research questions relates to a fundamental methodological issue, for which no simple answers exist: the vague character of the success criteria *usefulness*. To answer **RQ2** we will argue for usefulness through qualitative evidence, namely by providing some detailed analyses of chains in a real use-case. Automatically evaluating quality of chains is even more difficult. Remember that we want to use the extensional semantics of the concepts for determining the mappings, which makes the only comparison we have, namely an intensional gold-standard, difficult to justify. In our use-case, the line between what was identified to be a correct extensional mapping and what was an incorrect association was very fine, and studying this friction will be in our view an important future topic for this type of research.

Data, experiments and evaluation For our experiments we used 5 different ontologies from the Dutch election campaigns in 1994, 1998, 2002, 2003 and 2006. Our experiment were conducted by mapping each of ontologies with each other. Each ontology was used to annotate (around 5000) newspaper articles of the respective campaign. Some initial formal evaluation was done by comparing mappings with an existing manually created (intensional) alignment. Evaluating the quality of the chains is more tricky, as we will discuss in Section 5.3. The answer to **RQ2** therefore remains anecdotal, and finally rather unsatisfactory.

Applications and generality Capturing meaning shift, particularly the extensional associations, of concepts over time, can help communication scientists to apply analysis on the dynamics of the political developments over time. This line of research is also generalisable in many other areas where similar problems occur, such as development of medical systems, e-Science, knowledge management, and other social networks, *etc.*

2 Instance-based matching method

Instance-based ontology matching techniques have shown its capacity of dealing with matching cases where lexical and structural techniques could not be applied effectively [2, 3]. A straightforward method is to measure the common extension of concepts [4, 5]. The major limitation of this method is usually a lack of shared instances. Recently, we have investigated ways of detecting concept correlation using the similarity between their instances [2, 1]. In our case, coders used concepts to describe the content of newspaper articles. We consider an article as an instance of a concept, if the concept is used to describe this article. Our hypothesis is that, even if the ontologies during different election periods are different, two similar articles should have been coded using similar concepts. Therefore, finding similar instances can lead to similar concepts.

Let O_1 and O_2 be two ontologies which are used to annotate two instance sets I_1 and I_2. The instance-matching based method consists of two steps:

- Instance enrichment. For each instance i_1 in I_1, find the most similar instance j_2 in I_2. We consider i_1 to be an instance of the concepts which j_2 is described with. The same operation is applied in the other direction. In the end, an artificial common instance set is built.
- Concept matching. Each concept corresponds to a set of instances, including their real instances and those enriched in the previous step. A corrected Jaccard similarity measure is applied to calculate the similarity between concepts from different years. That is

$$\text{Jacc} = \frac{\sqrt{|{c_1}^i \cup {c_2}^i| * (|{c_1}^i \cup {c_2}^i| - 0.8)}}{|{c_1}^i \cup {c_2}^i|} \tag{1}$$

  where ${c_1}^i$, ${c_2}^i$ are the instance sets of two concept $c_1 (\in O_1)$ and $c_2 (\in O_2)$. [3]

Two concepts with sufficient similarity are considered mapped. A set of mappings between concepts of two ontologies form an alignment between the ontologies.

Instance matching There are different ways to match instances. A simple method is to consider instances as documents, and apply information retrieval techniques to retrieve similar instances (documents). We use a tf-idf weighting scheme which is often exploited in the vector space model for information retrieval and text mining [6]. The idea is that each document is represented by a vector, each element is a weight of a word which occurs in this document. Each word is weighted using its tf-idf value. Traditionally, a query is represented as a vector using the idf of the to-be-queried dataset. In our case, the same word is likely to have different importance in different datasets, therefore, while building the vector representation of each document, we use the corresponding idf values

[3] To avoid very high scores in the case of very few instances a 0.8 parameter was chosen so that concepts with a single (also shared) instance obtain the same score as concepts with, in the limit, infinitely many instances, 20% of which co-occur.

Year	Articles	Concepts used	Concept manually mapped to	
			existing concepts	added news concepts
1994	1502	101	54	37
1998	5635	312	154	135
2002	6323	370	201	110
2003	5053	299	190	89
2006	5126	580		

Table 1. Datasets and manual mappings of different years.

of words calculated within the dataset to which the document belongs. Based on such vectors, the cosine similarity is used to determine the similarity between two documents. In this way, instances from different datasets are matched, and the information is used for the enrichment process.

Chains of mappings After alignments are generated between multiple ontologies, with some ontologies involved in multiple alignments, it is possible to generate chains of mappings between a series of ontologies, in, for example, a chronological order.

Let A_{12} and A_{23} be the alignments between O_1 and O_2 and between O_2 and O_3. If there is a mapping in A_{12}, $< c_{1i}, c_{2j}, v_{ij}^{12} >$ and a mapping in A_{23}, $< c_{2j}, c_{3k}, v_{jk}^{23} >$, this results in a two-step chain of mapping from c_{1i} to c_{3k} via c_{2j}, with a confidence value $v_{ik} = v_{ij}^{12} \times v_{jk}^{23}$. When there are a series of alignments between O_1 and O_2, O_2 and O_3, until O_{n-1} and O_n, this will result in n-1-step chains of mappings

$$< c_{1i} \rightarrow c_{2j} \rightarrow \cdots \rightarrow c_{n-1,k} \rightarrow c_{nl}, v_{ij}^{12} \times \ldots v_{kl}^{n-1,n} > .$$

In this paper, we investigate 4 different kinds of mapping-chains, and investigate their usefulness in a practical application:

1. Top-1 forward chain (such as in Fig.:2): in each step, only the mapping with highest confidence is considered.
2. Top-n forward-chains (such as in Fig.:4): in each step, the the best n mappings are considered, starting with the first ontology.
3. Top-n backward-chains (such as in Fig.:5): in each step, the the best n are considered, starting with the last ontology.
4. Top-n kite (such as in Fig.:6): starting with the first ontology in each step, the mappings with the n highest confidence values for which there exist a top n mapping chain to the correct mapping (according to the gold standard).

3 Ontologies in Semantic Network Analysis

The series of ontologies to be mapped are the ontologies used to code newspapers during five recent Dutch elections taking place in 1994, 1998, 2002, 2003, and 2006. The articles were coded using the Network analysis of Evaluative Texts (NET) method [7], popularly used in the Semantic Network Analysis. During

each election year, annotators coded newspaper articles using the ontology available to them. Take as an example a sentence in a newspaper article during the election period in 2006.

Example 1. Het Openbaar Ministerie (OM) wil de komende vier jaar mensenhandel uitroeien. (*The Justice Department (OM) wants to eliminate human trafficking within the next four years.*)

The sentence is coded as `<om, -1,human trafficking>`, where `om` and `human trafficking` are two concepts in the ontology used in 2006, while `-1` indicates the Justice Department is negative about human trafficking. In this example, we consider this sentence to be an instance of the two concepts involved. All five ontologies are represented in the standard SKOS format [8]. Each concept has an prefLabel and possibly a few altLabel which are the synonyms of this concept and also used by coders in the coding process. Except the most recent election, all the newspapers are coded at the article level, but mainly based on the first three sentences. In 2006, the coding is at the sentence level.

The synonymous concepts were found manually. As shown in Table 1, the number of manually mapped concepts is smaller than that of concepts found in the actual coding. The reason is that new variations of the concepts were manually input to the database. These variations are very likely to be synonyms of concepts in the ontologies or pure typos, which were not covered during the manual mapping process.

Alignments between ontologies of previous years to the latest 2006 version have also been made manually. However, some concepts used in previous years cannot find the exact correspondences in the 2006 version. In that case, the domain experts added new concepts to the current version. The last column of Table 1 indicates how many new concepts were added during the manual aligning process. These new concepts are not used during the coding of 2006 articles, which means they do not have any instances in the 2006 corpus and were therefore not considered in our automated evaluation.

4 Base experiments: single mappings

Before focussing on chains of mappings, we need to show that our methods for calculating individual mappings are trustworthy. We first map all previous ontologies to the 2006 ontology. According to the extensional mapping technique, one concept can be related to multiple concepts, each with a certain amount of relatedness. As only one mapping for each concept was considered in the manual mapping results, for each concept, we take the mapping with the highest confidence value (*i.e.* , the corrected Jaccard similarity) as the final mapping of this concept. Table 2 shows precision and recall of these 1:1 mappings.

For all the concepts which were manually mapped with existing concepts in the 2006 ontology, we also measure the mean reciprocal rank (MRR) $mrr = \frac{1}{|C|}\sum_{i=1}^{C}\frac{1}{rank_i}$, where C is the set of concepts, the $rank_i$ is the rank of the concept which C_i should be mapped to. When C_i does not have a match, the

Year	Precision	Recall
1994	0.22	0.22
2002	0.36	0.35

Year	Precision	Recall
1998	0.33	0.33
2003	0.4	0.4

Table 2. Evaluation of 1:1 mappings based on the sentence and article level

Year	Concepts Found	Concepts matched	MRR
1994	532	39	0.41
2002	570	143	0.54

Year	Concepts Found	Concepts matched	MRR
1998	561	102	0.47
2003	570	136	0.58

Table 3. Mean Reciprocal Rank: from 2006 to the previous years, where "Concepts Found" is the number of concepts for which we have found some mappings, "Concepts matched" is the number of concepts for which we have recovered the correct mapping.

reciprocal rank is set to 0. A higher *mrr* indicates the correct matches are ranked in the more front position. Table 3 shows that the correct mapping is ranked on average within the top 10 proposed ones.

5 Main experiments: Chains of mappings

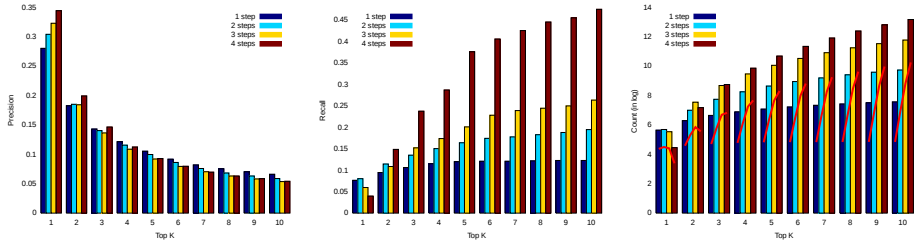
The main topic of this paper is to investigate the use of chains of mappings of concepts from ontologies from 1994 to 2006.

5.1 Quantitative analysis

Based on the manual alignments, we can measure the precision and recall of the chains. For each concept from 1994, top K mappings are taken into consideration, each of which will be expanded by its top K mappings too, and so on.[4] A chain can start from any concept of any year. For all chains with n steps, $n = 1, 2, 3, 4$, we measure the precision and recall respectively. A chain is considered to be correct if the two end-point concepts form a correct mapping according to gold standard, *i.e.* , the correctness is 1; a chain is partially correct and the correctness is the number of correct mappings on the way over the number of steps. In the end we take the average of the correctness of all individual chains as the final precision. The partial correctness is not considered when calculating the recall. The evaluation results are shown in Fig. 1.

Clearly, when considering more mapping candidates, more noisy data is included. Fig. 1 (c) gives the raw count of chains in terms of the choice of K. Note the numbers are in log scale. The red line indicates the number of chains with the two end-point concepts form a correct mapping judged by the gold standard.

[4] By expanding via different intermediate mappings from one concept, the total amount of chains is growing, but not exponentially. The reason is that concepts related to the starting concept tend to have a similar extensional semantics.



(a) Precision (b) Recall (c) Raw count

Fig. 1. Performance of the mapping chains. The x-axis indicates, for each concept, the top K related concepts are considered.

If only taking top 1, 2 or 3 related concepts, when the steps goes up, the number of correct chains drop, but the total amount of chains drops even faster. The precision of multi-step chains is actually higher than that of shorter chains. This suggests, even the absolute correctness of direct mappings may not be perfect, a multi-step chains of less perfect mappings may still lead to correct mappings over all time. Unfortunately, when taking more less related concepts, the number of correct chains goes up, but the sheer amount of total chains climbs up more rapidly, which cause the precision to drop in the end; the precision of multi-step chains is lower than that of shorter chains, that is the quality of the mappings degraded when the intermediate steps became longer.

5.2 Qualitative analysis of chains

A qualitative analysis of chains of mappings provides interesting insights in the value of the matching of ontologies over time for social sciences. A typical example of mapping of concepts over time will be discussed. The **domain-specific** political analysis will be supplemented by a methodological discussion (we call it **Metaanalysis** provided in italics) from the perspective of the usefulness of mapping-chains.

Let us start with an analysis of different chains for the two concepts asylum seekers ("asielzoekers") and senior citizens ("ouderen"). Figure 2 shows that the concept of asylum seekers (label: asielzoekers) is correctly mapped in each of the election years to the same concept at the highest rank. This result indicates that no or limited topic drift occurred. The confidence value slowly deteriorates, which might imply that the debate about asylum seekers has become more multi-faceted. An alternative explanation is that the number of concepts relating to asylum policy in the ontologies has increased because of the increasing political interest in asylum policy.

Fig. 2. Top 1 foward expansion for concept asielzoekers

Metaanalysis 1: *Even from a simple top 1 forward chain, some lessons can be drawn. However, note that analyses of this kind depend on the* ***confidence values****, which are rather* ***dubious*** *at best. Top 1 chains are more interesting objects of study when more drastic shifts occur.*

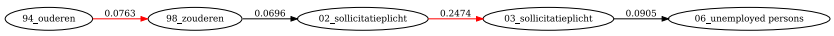


Fig. 3. Top 1 foward expansion for concept ouderen

Figure 3 shows that the elderly concept (ouderen) is only mapped to the expected concept between 1994 and 1998. The 1998 concept of the elderly is mapped to "obligation to apply for a job for the unemployed" (sollicitatieplicht). The abolition of the exemption of the obligation to apply for a job for the elderly was an election issues in 2002, which explains the link between these concepts. Both concepts should be considered as social security issues from a theoretical stance, since the elderly became an election issue during election campaigns with regard to special social security arrangement for the elderly. In 2003 the obligation to apply is correctly mapped to the 2002 concept. In 2006 the obligation to apply is mapped to the related concept of the unemployed.

Metaanalysis 2: ***Association versus similarity:*** *One of the crucial methodological problems is the formal meaning of the mappings between two concepts such as the elderly and the obligation to apply. Clearly, our instance-based methods find mappings with an extensional semantics,* i.e. *, the use of the concepts in annotating articles is related. A domain expert can identify elements of the intensional meaning that relate these two concepts in the specific case. Concretely, the issue "senior citizen" in 1998 and the issue "obligation to apply" in 2002 also share an intensional meaning. However, to the best of our knowledge there is no theory to formalise the relation between the extensional meaning and the intensional meaning of the mappings. In the following we will see examples where using association misses the goal of finding similarity, in particular when chains of mappings are considered.*

Although the asylum seeker concept is mapped correctly, the chains including top 2 concepts – represented in Figure 4 – give additional insights into the nature of the asylum debate over time. An analysis of the secondly ranked concepts shows with which issues asylum seekers have been associated during the election campaigns. In 1998 asylum seekers are mapped to the Cabinet of the time (kabinet kokmierlods), which follows from the fact that this Cabinet paid much attention to this issue. The second rank concepts in the following years indicate changes in the proposed policy measures. In 2002, when the anti-immigration party LPF came into power, asylum seekers are mapped to the constraint of the influx of refugees (instroom beperking). In 2003, after the LPF had left the government, it was mapped to a non-restrictive measure, assistance to illegal immigrants (opvang illegalen). In 2006 – the year in which another anti-immigration

Fig. 4. Top 2 foward expansion

party made an upsurge – was mapped to integration. It is worth noting that in 2006 the second rank mapping (integration) has a positive connotation, which suggests that the new anti-immigration party did not manage to influence public opinion to the degree the LPF managed in 2002.

> **Metaanalysis 3**: *This nice example shows a number of useful findings from the Communication Scientist perspective. The* ***chains of intensional meaning of the concepts at the time of their use provides interesting, and very subtle, developments for further investigation****. Given that the top 1 mappings are all correct, the interest lies particular in the second ranked mappings of each step, which tell a story of Dutch politics.*

With the expansion of the chain, some concepts that were mapped to secondly ranked concepts do not always directly relate to asylum seekers anymore. While the secondly ranked concept assistance to illegal immigrants in 2003 was plausibly mapped to illegal immigrants (illegalen), it was also mapped to organ donation (orgaandonatie). This latter mapping is explained by a particular political actor who propagated both the assistance to illegal immigrants in 2003 and organ donation in 2006.

The lower half of the figure does not directly relate to asylum seekers either, since the Cabinet in 2003 was not mapped at a high rank to asylum seekers anymore, but to the military aircraft Joint Strike Fighter (jsf) and business (bedrijfsleven), which in turn were mapped to concepts in the military and economical area. We omitted parts of the lower half for space reasons.

> **Metaanalysis 4**: *These are examples, where association can lead to unrelated concepts in 2 steps only. This highlights one of the biggest methodological challenges: how to distinguish useful and non-useful chains.* ***Early erroneous associations can turn large parts of the analysis practically useless.***

Figure 5 containing the backward chain starting from 2006, complements the information about the nature of the asylum debate. Studying the secondly ranked concepts mapped to asylum seekers in the backward direction, partially differing association with asylum seekers appear. Asylum seekers in 2006 are mapped to crime (criminaliteit) in 2003. In recent elections immigrants (including asylum seekers) have been regularly associated with crime by anti-immigration parties. Although these concepts are not directly related, they are related to each other

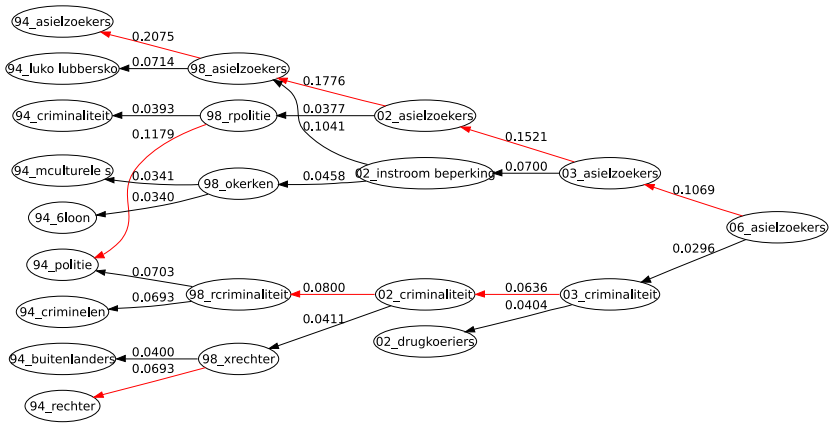


Fig. 5. Top 2 backward expansion

in the political reality. As in the forward mapping asylum seekers in 2003 is mapped to the constraint of the influx of refugees in 2002. In 1998 they are mapped to the police (rpolitie) and in 1994 to the Cabinet (luko lubbersko). The mapping to the police is in line with the mapping to crime in 2003.

> **Metaanalysis 5**: ***The chains of the concept mappings in two different directions are complementary***. *While it seems, for example, anomalous that churches (okerken) in 1998 is mapped to the constraint of the influx of asylum seekers in 2002, the mapping of the constraint of the influx of asylum seekers in 2002 to assistance to asylum seekers in 1998 in the opposite direction helps to explain this mapping, since churches played an important role in the assistance to asylum seekers.*

It is noticeable that the expansion chains do not expand exponentially, but still faster than we expected.[5] An interesting phenomenon is that mappings do not converge again, *i.e.* , once an association happens in one year, it usually does not associate back the following year to the same topic. This is an interesting finding, for which we do not have an explanation.

> **Metaanalysis 6**: ***The expansion factor is meaningful*** *in two ways: it gives an indication on the debate itself, but it can also be an indication for the mapping quality. The smaller the tree is, the more closely related are the associated concepts, which might indicate that the mapping quality is better than that for larger trees.*

The kite with two correct endpoint concepts integrates the information from the previous figures. In Figure 6, it becomes clear that the concept of asylum seekers is correctly mapped from one election year to another. Additionally it shows that the asylum debate is both associated with central concepts in the

[5] In the top 2 chains we studied the average width was 12.

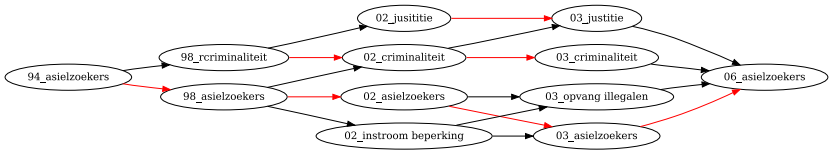


Fig. 6. Kite with two correct endpoint concepts

asylum policy debate, the constraint of the influx of refugees and the assistance to illegal immigrants, and to crime, a concept that is, however not directly related to asylum seekers, related to the concept in the political reality.

> **Metaanalysis 7**: ***Kites from end to end seem the most useful way of illustrating concept drift****, as incorrect associations are eliminated (because they are not mapped back to the start-concept). However, such a kite is less fine-grained than the forward chain: the subtle change between negative and positive connotation between 02_instroom beperking and 06_integration is lost.*

5.3 Discussion

The issues discussed in the previous section mostly center around two questions: the usefulness of our proposed mapping chains, and the meaning of the extensional mappings in the first place. Studying a representative sample of chains through our domain expert indicates clearly that mapping chains can be interesting given the reasonable quality of the individual mappings we can produce.

However, apart from the concern about the accuracy of the mapping method used here, two problems are apparent to which we do not yet have a satisfactory solution: the lack of a notion of correctness of extensional mappings in the chains, and the evaluation of this correctness. Our manual analysis shows plenty of examples where an associative semantics of mappings based on the extensions of concepts corresponds to an intensional relation of the meaning between of two concepts. However, in other examples these associations often totally diverge from what domain experts find acceptable intensional similarities. We have no intuition yet how to address this problem, *i.e.* , how to formalise and study it. The most promising solution for addressing this problem is to consider the kite structures, in which the disambiguation of mappings is achieved by requiring a mapping back to the original concept. In that way wild mismatches can be eliminated from the temporal chains. Many interesting parts of the temporal "story" get lost in this approach, though.

The second problem is the strongly related problem of evaluation: so far we found only two ways of evaluating our constructs: 1) comparing the chains with an intensional gold standard, and 2) having a domain expert evaluate each of the chains. Obviously, the first option is methodologically not valid, as we evaluate against something we know not to be the solution. The second approach is more acceptable from the domain perspective, but manual checking is very expensive

and difficult to quantify, if it is not yet impossible to identify all "intertesting chains." The only practical solutions we found so far is to use the spreading factor of the top K chains over time. In our view, the fewer leaves such a tree has, the semantically closer the mappings should be. However, this idea is build on intuition rather than empirical findings.

6 Conclusion

In this paper we introduced different representations of mapping chains and evaluated them over sequences of political ontologies in a Media study driven by Communication Scientists. We used instance-based mappings between pairs of ontologies to calculate sequences of extensional mappings and show, for a usecase analysing Dutch election campaigns, some interesting qualitative findings.

Apart from these stimulating examples we also provided an initial evaluation of our proposal, both qualitative and quantitative. However, this evaluation is tricky, as neither the notion of correctness of an extensional mapping is well-defined, nor do we have a sound evaluation methodology yet.

For us the general lessons for the ontology mapping community is twofold: that the semantics of mappings is not yet fully understood, particularly, *w.r.t.*, extensional semantics, and that mappings in a dynamic context are challenging, and worthwhile, objects of study in addition to their known static variants.

References

1. Schopman, B., Wang, S., Schlobach, S.: Deriving concept mappings through instance mappings. In: Proceedings of the 3rd Asian Semantic Web Conference, Bangkok, Thailand (2008)
2. Wang, S., Englebienne, G., Schlobach, S.: Learning concept mappings from instance similarity. In: Proceedings of the 7th International Semantic Web Conference (ISWC 2008). Volume 5318 of Lecture Notes in Computer Science., Karlsruhe, Germany, Springer (October 2008) 339–355
3. Wang, S., Isaac, A., Schopman, B., Schlobach, S., van der Meij, L.: Matching multilingual subject vocabularies. In: Proceedings of the 13th European Conference on Digital Libraries (ECDL2009), Corfu, Greece (September 2009)
4. Euzenat, J., Shvaiko, P.: Ontology Matching. Springer Verlag (2007)
5. Isaac, A., van der Meij, L., Schlobach, S., Wang, S.: An empirical study of instance-based ontology matching. In: Proceedings of the 6th International Semantic Web Conference (ISWC 2007). Volume 4825 of Lecture Notes in Computer Science., Busan, Korea, Springer (2007)
6. Salton, G., McGill, M.J.: Introduction to Modern Information Retrieval. McGraw-Hill (1983)
7. Van Cuilenburg, J., Kleinnijenhuis, J., De ridder, J.: Towards a graph theory of journalistic texts. European Journal of Communication **1** (1986) 65–96
8. Isaac, A., Summers, E.: SKOS Primer. W3C Group Note (2009)