

# SPARQLing Conversation: Automating The Bechdel-Wallace Test

Dr. K. Faith Lawrence  
Royal Irish Academy  
kf03r@ecs.soton.ac.uk

## Keywords

Narrative, Semantic Web, Ontology, TEI, RDF, SPARQL

## 1. INTRODUCTION

One of the goals in digital narrative is to describe the elements within a story in such a way that they can be analysed and queried. In this paper we present one attempt at representing and applying a well-known rubric, the Bechdel-Wallace test, to a work of fiction. While the final results were disappointing we describe the process that we undertook and highlight the areas of difficulty and complexity which we found and which we will need to address in future work.

## 2. LINKED DATA FROM TEI

Jewell[2] presented a methodology for the automated creation of RDF from TEI encoded scripts. This procedure allowed us to generate basic descriptions of the events that occur within a narrative including location, involved characters and referenced characters. Although in its initial stages this work has created a resource which can be accessed via a SPARQL endpoint or via an editing interface which is currently under development (see figure 1). Two test datasets were used for the purposes of the work detailed below - a small dataset which was manually created to include example cases and the larger one which represented the narrative storyline of A Midsummer Night's Dream[6]. This second dataset was automatically generated from the TEI encoded version of the play held at the Perseus Digital Library.

## 3. THE BECHDEL-WALLACE TEST

The Bechdel-Wallace (B-W) test was popularised in an 1985 online comic strip 'The Rule'[1] in which a character claimed never to watch a film unless it passed three criteria which she listed as:

1. It has to have at least two women in it,



Figure 1: Screenshot of the Event Viewer page of the Shakespeare Editor

2. Who talk to each other,
3. About something other than a man

While not passing judgement on the quality of an individual work being assessed, this rule has been seen as an interesting thinking-point both in regards to single pieces and for looking at, and raising awareness of, trends of representation across bodies of works. A frequently used variation takes this rule even further to require that the female characters in question are named i.e. that the characters have a name and are not just referred to by their description or job title in the credits.

The OntoMedia ontology[3] was developed to allow the description of narrative content in such a way as to allow this type of query, and the Bechdel test seemed an ideal test scenario for our data. Since the TEI encoding of the Shakespeare play did not include the gender of the characters and the RDF editor interface was not, at the time of writing, able to allow descriptive expansions, this information was added manually. Other than these changes the data was not corrected and the ways in which the roughness of the data affected the results of our queries will be discussed below.

## 4. TALKING ABOUT SOMETHING OTHER THAN A MEN

The automated RDF creation tool that was used to generate the data interprets the TEI speech tags as representing an

ome:Event in which the speaker (as indicated by the original TEI) is the primary entity within the event. The tool then assumed that every other character who was calculated to be in the same location as the speaker was involved in that particular ome:Event on the basis that they were deemed to be within hearing distance. It was recognised that this was not universally valid but it was felt that this offered a reasonable default position given the information that was available via the encoding. Within the OntoMedia ontology conversations are generally classed as social events and with no further details guaranteed the broadest categorisation was used rendering all such events of the class ome:Social.

Using this as a premise we can construct a simple query based on the first two requirements of the B-W rule, complete with named-characters variation:

```
SELECT DISTINCT ?event ?label ?name1 ?name2
WHERE {
?event a <http://purl.org/ontomedia/core/
expression#Social>;
<http://purl.org/ontomedia/core/
expression#has-subject-entity> ?
char1 ;
<http://purl.org/ontomedia/core/
expression#involves> ?char2 ;
rdfs:label ?label .
?char1 <http://purl.org/ontomedia/ext/
common/trait#has-trait> <http://
contextus.net/resource/meta/Female>;
<http://xmlns.com/foaf/0.1/name> ?name1
.
?char2 <http://purl.org/ontomedia/ext/
common/trait#has-trait> <http://
contextus.net/resource/meta/Female>;
<http://xmlns.com/foaf/0.1/name> ?name2
.
FILTER (?char1 != ?char2) .
}
```

One noted lack in SPARQL 1.0 is an simple way to carry out queries containing negation. The simplest way to implement bound negation is through use of the OPTIONAL command (see below).

```
SELECT ?event ?label ?name1 ?name2 WHERE {
?event a <http://purl.org/ontomedia/core/
expression#Social>;
<http://purl.org/ontomedia/core/
expression#has-subject-entity> ?
char1 ;
<http://purl.org/ontomedia/core/
expression#involves> ?char2 ;
rdfs:label ?label .
?char1 <http://purl.org/ontomedia/ext/
common/trait#has-trait> <http://
contextus.net/resource/meta/Female>;
<http://xmlns.com/foaf/0.1/name> ?name1
.
?char2 <http://purl.org/ontomedia/ext/
common/trait#has-trait> <http://
contextus.net/resource/meta/Female>;
<http://xmlns.com/foaf/0.1/name> ?name2
.
OPTIONAL { ?event <http://purl.org/
```

```
ontomedia/core/expression#refers-to> ?
char3 .
?char3 <http://purl.org/ontomedia/ext/
common/trait#has-trait> ?gender . }
FILTER (?char1 != ?char2 && (!bound(?gender
) || ?gender != <http://contextus.net/
resource/meta/Male>)) .
} ORDER BY ?event
```

This query was successfully run on the test data but when it was applied to Midsummer Night's Dream it was quickly apparent that it was not returning the desired results because in those situations where multiple characters were referred to within the speech then the event was returned as long as at least one of the characters referred to did not have the trait 'Male'.

Event	References	Gender	Returned	Desired
13	-	-	Yes	Yes
40	Helena	Female	Yes	Yes
171	Hermia	Female	Yes	No
171	Lysander	Male	No	No
304	Demetrius	Male	No	No
304	Lysander	Male	No	No
336	Lysander	Male	No	No

Queries of this type are notoriously problematic in SPARQL 1.0 but it is hoped that with the advent of increased support for the extended negation features in SPARQL 1.1 we can revisit this question. It was also noted that a number of the events that were returned should not have been, since to a human eye they clearly reference a male character but not by name. For example, in event 13 Hermia says "I would my father look'd but with my eyes" (Act 1: Scene 1) but the automated process did not identify the reference 'father' with the character of Egeus. A more sophisticated analysis might be able to identify and encode these indirect references thus improving the accuracy of the system. The intention was that the automated description would be superseded by a corrected version of events and the supplementation of the base information with missed references was one aspect of this error checking procedure. It is in cases such as this that the results would be immediately valuable.

In the rest of this paper we will consider some of the other issues that were raised during the investigation.

## 5. REFINING THE QUESTION

### 5.1 What is a Conversation?

There is debate among those who apply the Bechdel-Wallace test as to how rigorously the rule should be applied. If a conversation has multiple topics, as a long exchange might well do, then does a reference to a man at any point within the conversation invalidate the entire exchange? Or can one section of the conversation be regarded as meeting the requirements and therefore validate the exchange even if other parts do not satisfy the criteria? In the latter case, how do we define the start and end of a conversational exchange? The given understanding is that the test is intended to be a low barrier and, on that basis, we will take the minimum possible definition, i.e. a single exchange, and regard that as all that is needed for the work to pass.

Entities are deemed to have three levels of inclusion in events: the subject (the main/active entity), the object (secondary entity) and at the most generic level any other character who is in any way part of the event is defined as involved. This latter may represent characters who are merely witnesses to the event occurrence. Since the automation system was not able to divine from the TEI which of the potential listeners was being addressed at any given time it was not possible to promote any entity, or group or entities, to being status of event object. Ideally the initial component of the query would reflect that a conversation between characters would be modelled as occurring between the subject and object entities <sup>1</sup>

```
?event a <http://purl.org/ontomedia/core/
  expression#Social>;
  <http://purl.org/ontomedia/core/
    expression#has-subject-entity> ?
    char1 ;
  <http://purl.org/ontomedia/core/
    expression#has-object-entity> ?
    char2 ;
  rdfs:label ?label .
```

In this case a single event could be held to fulfill the necessary criteria and its existence would indicate that the work passed the test. Lacking this specificity, we must question how we model the concept of talking to someone. We would argue that a character can be understood to be part of the conversation rather than just witnessing it if they respond. This can be represented by expanding the query to require an exchange of speech between two given female characters where neither part of the exchange refers to a male character.

```
SELECT DISTINCT ?event ?label ?name1 ?name2
  WHERE {
?event a <http://purl.org/ontomedia/core/
  expression#Social>;
  <http://purl.org/ontomedia/core/
    expression#has-subject-entity> ?
    char1 ;
  <http://purl.org/ontomedia/core/
    expression#involves> ?char2 ;
  rdfs:label ?label .
OPTIONAL { ?event ?link ?event2 .
  ?event2 a <http://purl.org/ontomedia/
    core/expression#Social>;
  <http://purl.org/ontomedia/core/
    expression#has-subject-entity> ?
    char2 ;
  <http://purl.org/ontomedia/core/
    expression#involves> ?char1 ;
  rdfs:label ?label2 . }
...
FILTER (bound(?event2) && (?link = <http://
  purl.org/ontomedia/core/expression#
  precedes> || ?link = <http://purl.org/
  ontomedia/core/expression#follows>)).
} ORDER BY ?event
```

The final stage of this query refinement would be to allow for either pattern to matched as valid and therefore return

<sup>1</sup>For the sake of simplicity I am presenting the case where subject and object are single entities rather than groups. This latter case will also have to be addressed.

a result set comprised of the union of any matching single event where the object is defined and any matching paired events in which the speaker is either replying to or replied to by a female character who is otherwise involved in the event.

## 5.2 Female? Male? Moving Beyond Binary Gender Classifications

In the examples above we have used two of the four top-level gender classes, `omt:Male` and `omt:Female`, defined in the OntoMedia ontology. The OntoMedia ontology differentiates between **Sex** (the biological categorisation) and **Gender** (the social categorisation). This decision was taken to allow the description of situations where these two traits were not the same, however it maintains a relationship between the two through stating the gender identity that an instance of a given biological or psychological classification of sex is associated with. In this way it is possible to build up a much more complex model of an entity's gender and to link, compare and contrast the potentially different 'sexes' associated with a given character via the genders those attributes are identified with. However it is equally important to recognise the difference between a character cross-dressing, impersonating another gender or identifying with another gender and how this is presented within the fictional universe and beyond the bounds of its contextual universe to the audience.

With fiction we are often given a very surface representation of a character and so the default level of information that we as the audience are given are at the gender rather than the biological level. However some fictional narratives deliberately play on or reference this difference, from Shakespeare's superficial cross-dressing farces to serious dramas such as 'Transamerica'[7] or, more tangentially, 'Let The Right One In'[4] or 'Inception'[5]. In these examples, the Shakespearean characters may change their clothes and the gender that they project to other characters within the context of the universe but their gender identity remains unchanged. And while Bree (Transamerica) and Eli (Let The Right One In) should clearly be identified as female for the purposes of the test, it would be harder to argue that Eames (Inception) is ever anything other than male as his periods of time as a woman are presented as gender impersonation rather than as shift in gender identity.

While the first criteria in the Bechdel-Wallace test is that the work contains two women we would argue that this was intended as a short hand way of saying "two characters who are not men". However it is important to note that "[the work] has to have at least two women in it" and "[the work] has to have at least two !(men) in it" are not the same statement since the set !(men) is much larger than the set (women). We raise the issues above to show the complexities that need to be addressed and would argue that the first step in doing so would be to alter the gender identifier from specifying a female character to negating a male one.

```
...
?char1 <http://purl.org/ontomedia/ext/
  common/trait#has-trait> ?gender;
  <http://xmlns.com/foaf/0.1/name> ?name1
.
?char2 <http://purl.org/ontomedia/ext/
```

```
common/trait#has-trait > ?gender2;  
<http://xmlns.com/foaf/0.1/name> ?name2
```

```
...  
FILTER (?char1 != ?char2 && (!bound(?gender  
  ) || ?gender != <http://contextus.net/  
  resource/meta/Male>) && (!bound(?  
  gender2) || ?gender2 != <http://  
  contextus.net/resource/meta/Male>)) .
```

## 6. CONCLUSION

The work presented in this paper created more questions than results however these questions are ones that need to be considered. Our initial hope was to create a simple query which could be applied to any narrative description that used the OntoMedia ontology and discover immediately whether it passed the Bechdel-Wallace test and which specific sections of dialogue were responsible for that result. Instead we have raised a number of valuable issues both in how we interpret the questions that we want to ask and how we then formulate them.

## References

- [1] A. Bechdel. *Dykes to Watch Out For*, page 22. Firebrand Books, 1986.
- [2] M. O. Jewell. Semantic screenplays: Preparing tei for linked data. In *Digital Humanities 2010*, June 2010. Paper presented as part of panel: Scanning Between the Lines: The Search for the Semantic Story.
- [3] K. F. Lawrence. *The Web of Community Trust - Amateur Fiction Online: A Case Study in Community Focused Design for the Semantic Web*. PhD thesis, Electronics and Computer Science, University of Southampton, 2007.
- [4] J. A. Lindqvist. *Let The Right One In*. Film.
- [5] C. Nolan. *Inception*. Film, 2010.
- [6] W. Shakespeare. *A Midsummer Night's Dream*. Thomas Fisher, 1600.
- [7] D. Tucker. *Transamerica*. Film, 2005.