# BOLT IR Thematic Grouping Assessment Guidelines

## Version 1.0, December 20, 2013
## Linguistic Data Consortium

1. Introduction ..... 3
2. Context and Motivation ..... 3
3. BOLT IR Decision Tree ..... 4
4. Rules of Thumb ..... 6
5. BOLT IR Queries, Views, Groups and Citations ..... 8
5.1 The Query ..... 8
5.2 Views, Groups and Labels ..... 8
5.3 The Citation ..... 9
5.4 Views, Groups and Citations - Examples ..... 9
6. Assessment Procedure Introduction ..... 11
7. Group-Level Assessment ..... 12
7.1 How cohesive is this group? ..... 12
7.2 Does this group represent a useful/appropriate way to cluster citations for this query? ..... 13
7.3 Does the group label reasonably capture what these citations have in common? ..... 14
7.4 Provide an improved group label ..... 15
8. Citation-Level Assessment ..... 16
9. View-Level Questions ..... 16
9.1 Does this view represent a useful/appropriate way to cluster groups for this query? ..... 17
9.2 Does the view label reasonably capture what these groups have in common? 17
9.3 Provide an improved view label ..... 18
10. Conclusion ..... Error! Bookmark not defined.

## 1. Introduction

The BOLT (Broad Operational Language Translation) Program will create new techniques for automated translation and linguistic analysis for informal, conversational genres in Egyptian Arabic, Chinese and English.

In BOLT Information Retrieval (IR) task, systems are required to process natural language queries in English and return relevant answers from a large corpus of Egyptian Arabic, Chinese and English discussion forum threads harvested from the web. Systems must return a list of responses (known as citations) for each query, translating responses into English where required. Human assessors then review citations for correctness.

After relevance assessment, systems perform what is known as *thematic grouping*. The goal of thematic grouping is to minimize redundancy in query results presented to an end user. In this task, systems automatically cluster correct citations into **views** containing one or more **groups** of related citations. Systems may produce up to three views per query, and each view may contain at most 30 groups with no more than 100 citations each. Citations may belong to more than one group per view.

This document describes the process for assessing thematic groupings in the BOLT Phase 2 IR Evaluation[1].

## 2. Context and Motivation

To help you understand the motivation for this task, imagine a busy monolingual English researcher who has a complex question. This researcher has access to a large, multilingual collection of documents where the answers can be found. Because the researcher is busy, there isn't time to manually look at all the documents in the collection. And because the research is a monolingual English speaker, it wouldn't be possible to find all the answers even if there was plenty of time.

In the primary relevance task, the researcher types the question into the BOLT IR machine, and gets back a list of short, relevant answers drawn from the document collection. Answers are automatically translated into English, and contain pointers to the original document so the researcher can follow up if desired. The relevance assessor's job is to rate the quality of the answers returned by the BOLT IR machine.

In the current thematic grouping task, the researcher types the question into the BOLT IR machine, and gets back a list of short, relevant answers drawn from the document collection that are grouped into a thematically-related clusters. The grouping assessor's job is to assess the quality of the clusters provided by the BOLT system, along several dimensions: the internal cohesiveness of the group, the

[1] For BOLT Phase 2 the thematic grouping task is a pilot evaluation limited to English.

appropriateness of the group/view to its query, whether a given citation is an appropriate member of its group, and the suitability of the group/view labels provided by the BOLT system.

## 3. BOLT IR Decision Tree

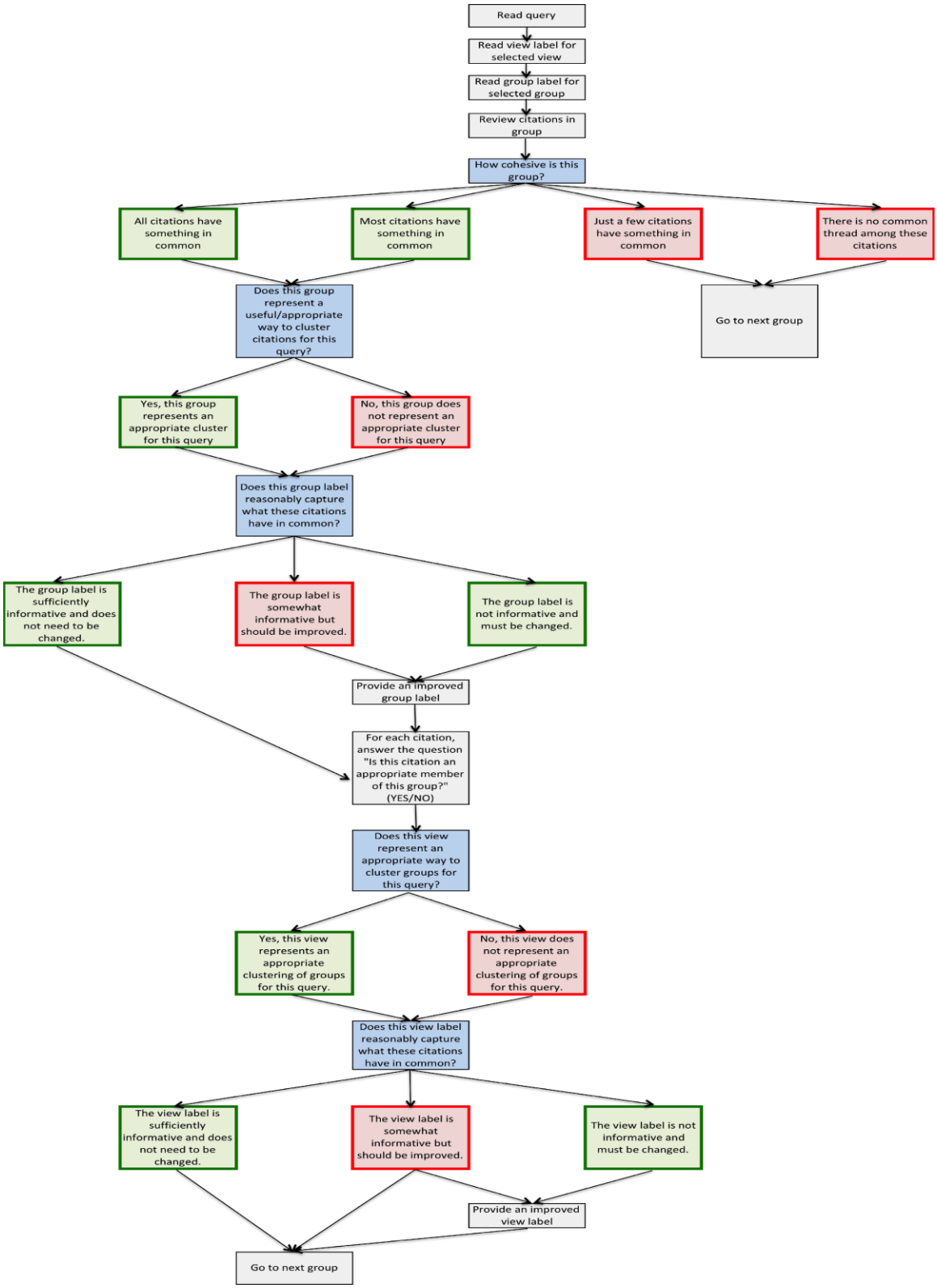
Thematic grouping assessment follows a decision tree. A decision tree depicts each question to be answered along with all possible outcomes and implications.

The BOLT IR grouping assessment annotation tool has been designed to support this decision tree model. You will be assigned a "kit" containing all of the groups and citations for a single view of a query. For kit, you will start by answering questions about a single group within that view. Once you have answered questions about the group, you will judge each citation within that group. You will then continue this process until all groups in the view have been fully assessed. After you have addressed all the groups in the view, you will answer some questions about the view itself. Then you will mark the kit Done and move onto the next kit.

In a decision tree, your answers to one question will determine future questions for that group or view. For instance, if you judge that there is no common thread among the citations in a group, you will not be asked further questions about that group, nor will you be asked to judge whether the citations in that group properly belong to it.

The complete decision tree is shown on the following page. The decision tree is intended to give you an overview of the decision making process for this task. It is not necessary to memorize the decision tree. The annotation tool has been designed to present questions to you one at a time, with all of the logical dependencies applied automatically.

We will examine each question in turn, in the sections that follow.




**BOLT Phase 2 - IR Evaluation Thematic Grouping Assessment Decision Tree**

## 4. Rules of Thumb

The decisions you must make for this task have some degree of subjectivity. In order to increase consistency among different annotators, you should follow these rules of thumb when judging thematic groupings.

### Rule 1: Be reasonable

Do not be excessively literal or overly strict when judging groupings.

**Example**

**Query:** xxx

**View Label:**
XXX

**Group Label:**
XXX

**Citation:**
XXX

Discussion

### Rule 2: Be generous

If you are on the fence about how to answer a question, give the machine the benefit of the doubt. This means that you should answer questions as follows when you can't decide:

**Question: How cohesive is this group?**

**Possible answers:**

a) All citations have something in common
b) Most citations have something in common
c) A few citations have something in common
d) There is no common thread among these citations

→ If you cannot decide between b) and c), choose *b) Most citations have something in common.*

**Question: Does this group represent a useful/appropriate way to cluster citations for this query?**

**Possible answers:**

a) Yes, this group represents an appropriate cluster for this query.
b) No, this group does not represent an appropriate cluster for this query

→ If you cannot decide between these options, choose *a) Yes, this group represents an appropriate cluster for this query.*

**Question: Does this group label reasonably capture what these citations have in common?**

**Possible answers:**

a) The group label is sufficiently informative and does not need to be changed.
b) The group label is somewhat informative but should be improved.
c) The group label is not informative and must be changed.

→ If you cannot decide between a) and b), choose *a) The group label is sufficiently informative and does not need to be changed.*
→ If you cannot decide between b) and c), choose *b) The group label is somewhat informative but should be improved.*

**Question: Is this citation an appropriate member of this group?**

**Possible answers:**

a) YES, the citation belongs in the group.
b) NO, the citation does not belong in the group.

→ If you cannot decide between a) and b), choose *a) YES, the citation belongs in the group..*

**Question: Does this view represent a useful/appropriate way to cluster groups for this query?**

**Possible answers:**

c) Yes, this view represents an appropriate clustering of groups for this query.
d) No, this view does not represent an appropriate clustering of groups for this query

→ If you cannot decide between these options, choose *a) Yes, this view represents an appropriate clustering of groups for this query.*

**Question: Does this view label reasonably capture what these groups have in common?**

**Possible answers:**

d) The view label is sufficiently informative and does not need to be changed.
e) The view label is somewhat informative but should be improved.
f) The view label is not informative and must be changed.

→ If you cannot decide between a) and b), choose *a) The view label is sufficiently informative and does not need to be changed.*
→ If you cannot decide between b) and c), choose *b) The view label is somewhat informative but should be improved.*

## 5. BOLT IR Queries, Views, Groups and Citations

### 5.1 The Query

Before you begin answering any questions you must read the natural language query in the info panel at the top of the annotation tool. Make sure you understand the query before you proceed.

You will be assessing thematic groupings for many different kinds of queries. Queries generally ask for information about persons, organizations, locations, facilities, events, movements, practices or customs, products, publications, laws, awards, diseases, or abstract entities. All queries are written in natural language, as a single English sentence.

Examples:

```
Query: Describe Russia's relations with the nation of Georgia.

Query: Why do people to leave the Church of Latter Day Saints?

Query: What are the effects of outsourcing jobs?
```

**NOTE:** If you don't understand the query after reading it, consult your team leader right away.

### 5.2 Views, Groups and Labels

A view is the top-most grouping for a query and is a container for more fine-grained groupings of citations. For each query, up to three different views are supported. A view is intended to be a useful way to organize citations for a user. For instance, in a query asking *What do people think about Pope Shenouda III?,* View 1 might contain groups of citations organized by different GeoPolitical entities, while View 2 might contain groups of citations organized by different Person entities.

Each view is automatically labeled by the BOLT IR System. View labels are not required to be informative, though they may be in some cases. For instance, in the *Pope Shenouda III* example, View 1 might be labeled *Groups displayed by: GPE*, or it might be labeled *x_m0.tf_2.euc0*. Because view labels are not required to be informative, it is very important that you ignore the view label when answering questions about the appropriateness of the view to its query. You will be asked to assess the suitability of the view label itself in a separate question.

Within each view you will be presented with up to 30 groups. In the *Pope Shenouda III* example discussed above, View 1 contains groups of citations organized by different GeoPolitical entities. So Group 1 within this view might contain citations discussing what people in Egypt think about Pope Shenouda. Group 2 might contain

citations discussing what people in Cairo think about Pope Shenouda, while Group 3 focuses on Alexandria and Group 4 on Abbasiya. Keep in mind that citations can belong to more than one group. So in this example, you might see the same citation in Group 1 (Egypt) and Group 2 (Cairo), since Cairo is in Egypt. As with view labels, group labels are not required to be informative.

BOLT IR Systems are also permitted to output one Miscellanous group for each view which can serve as a catchall for citations that don't fit well into other groups. Citations in the Miscellanous group are not intended to be related to or redundant with each other. Miscellanous groups are not assessed.

There is no prior expectation about the level of granularity for views and groups and the granularity may vary from one query to the next. However, because they are contained within views, groups are generally expected to be more fine-grained than views.

### 5.3 The Citation

Citations are answers to the query, automatically generated by the BOLT IR systems. All the citations in the groups/views you judge have already been assessed as relevant to the query. For this task, citations are always in English and are drawn from English source documents; in other words they are not machine translations from a foreign language into English.

Citations are always short – no more than 250 characters. This means that citations are sometimes missing context that you will need to accurately assess thematic groupings. In this case, you will need to check the original source document to gain a complete understanding of the citation's meaning.

Some examples of citations:

> **Citation:** I've already said that I think outdoor smoking bans are silly and unnecessary, I also think any kind of smoking ban on private vehicles or homes is intrusive and unenforceable.
>
> **Citation:** Most tourists enjoy the country from the comfort of their tour buses but on the streets it's a different story.
>
> **Citation:** …
>
> **Citation:** …

As you can see, some citations are very clear even without additional context, while others require additional context to understand their meaning.

### 5.4 Views, Groups and Citations - Examples

Below are some examples of views and groups with their labels, along with some citations that appear in each group.

| Query | View Label | Group Label | Citations Within This Group |
|---|---|---|---|
| **What are people saying about returning adopted foreign children to their home countries?** | **Groups displayed by: GPE** | ***Citations related to: Chinese*** | We predicted that skilled foreign workers would increasingly get fed up and return to countries like India and China where the economies were booming. |
| | | | The first year of a babies life is the most important medically, and the countries that have the highest rate of adoption (traditionally Russia, Romania, Guatemala, and China) have the lousiest health care system. |
| | | | But human resources directors in India and China told us that what was a trickle of returnees a decade ago had become a flood. |
| | | | At the end of this month, they are going to China where they've arranged to adopt a little one-year-old boy. |
| | | | South Korea should call on China to ensure that all children go to school regardless of their legal status, Pearson said. |
| | | | Under international law, and its own domestic laws, China is legally obligated to grant all children access to free elementary education, regardless of their nationality or legal status. |
| | | ***Citations related to: U.S.-Russian*** | " Parents enter into it ( foreign adoption ) with positive motivations but , in a sense , they are a little bit blindsided by their desire to adopt , " said LaBarbera , who specializes in the psychological evaluation of children and has worked with a of . |
| | | | I urge everyone who has adopted from a foriegn country to also sponsor at least one child that hasn't been adopted. |
| | | | The first year of a babies life is the most important medically, and the countries that have the highest rate of adoption (traditionally Russia, Romania, Guatemala, and China) have the lousiest health care system. |
| | | | Last year, nearly 1,600 Russian children were adopted in the United States, and more than 60,000 Russian orphans have been successfully adopted there, according to the National Council For Adoption, a U.S. adoption advocacy nonprofit group. |
| | | | She said she and her daughter went to Russia together to adopt the boy, and she believes information about his behavioral problems was withheld from her daughter. |
| | | | "The Russian orphanage officials completely lied to her because they wanted to get rid of him," Nancy Hansen said. |
| | | | |
| | | | |

| Query | View Label | Group Label | Citations Within This Group |
|---|---|---|---|
| **Did Obama deserve to win a Nobel Prize?** | **x_m0.tf_2.euc0** | **president__barack__days** | Obama won a Nobel Prize based on his first twelve days in office. |
| | | | In a decision as shocking as Friday's surprise peace prize win, President Obama failed to win the Nobel Memorial Prize in Economic Sciences Monday. |
| | | | Our President is a Laughingstock: Obama Awarded Nobel Peace Prize |
| | | | Obama reads a couple speeches and is elected president just twelve days when his name is put in the running and voila--even though there are no accomplishments--the Nobel Peace Prize. |
| | | | Whats amazing to me, is how those on the right are blaming Obama for being awarded the prize, something he had no control over, and was as surprised as the rest of us, to hear he had won it. |
| | | | I enjoy the fact that our healthcare system will forever be known as "Obamacare" I enjoy how pissed off the rightwing is that Obama won a Nobel Peace Price |
| | | **bush__award__president** | Fairly gentle satire of the Nobel Peace Prize award to President Obama, although they do get to the only reason for the award: Obama is not George Bush. |
| | | **deserves__win__peace** | The Nobel peace Prize sometimes has dubious recipients (the vast majority do deserve it). |
| | | | Obama quips his Nobel prize wasn't deserved - The Oval: Tracking the Obama presidency |
| | | | Even under those criteria, obama hardly deserves a "peace prize". |
| | | | Of course when Obama is winning a Nobel peace prize for NOTHING then the prize is meaningless. |
| | | | BTW there is a list of the far more deserving people Obama won over this year: |

## 6. Assessment Procedure Introduction

For the thematic grouping assessment task, kits consist of a single query with up to 3 views, 30 groups per view, and 100 citations per group.

Prior to answering questions about groups, citations and views you must first read the Query and be sure you understand it. The Query is shown in the info panel at the top of the annotation GUI. Also look at the View Label, shown directly under the Query text. Keep in mind that the View Label is not required to be informative.

## 7. Group-Level Assessment

After reviewing the Query, View Label and other preliminary information, you are ready to answer the questions pertaining to the first group in the kit. The annotation GUI will automatically select the correct group for you to judge.

### 7.1 How cohesive is this group?

The first group question you must answer is *How cohesive is this group?* There are four possible answers:

a) All citations have something in common
b) Most citations have something in common
c) A few citations have something in common
d) There is no common thread among these citations

To answer this question, you need to look at the citations contained within the group and see if you can find a common thread. The group and view labels might provide you with some hints, but it is important to keep in mind that the labels are not required to be informative or accurate. You should not assume that because the label is not informative, the group is not cohesive. It may happen that the group label is quite descriptive but the citations contained within the group have nothing in common. Similarly, the citations may have something in common that isn't reflected in the label. Therefore, you should not let the group label bias your answer to this question; the question should be answered on the basis of the citations that comprise the group.

As described in Section 5.3, citations are necessarily short. It may therefore be necessary for you to see one or more citations in context by consulting the source document(s) before you can determine whether the citations in a group have anything in common. For instance, a BOLT IR System may create a group of citations containing statements from Chinese leaders, but the citations themselves do not contain contextual information about who is speaking, so you would not detect the common thread unless you checked the source document(s).

As you are reviewing citations within this group, you might want to use the Scratchpad on the lefthand panel to keep notes about what these citations have in common. Assuming you have a theme in mind, you might also want to keep track of whether or not each citation conforms to that theme. You can do this by clicking the Yes/No radio buttons in the middle Citation panel. Keep in mind that your understanding of the common thread or theme may evolve as you review additional citations and so the Yes/No judgment for any individual citation may also evolve. Therefore, you cannot "finalize" your citation judgments at this stage of assessment.

After you finish reviewing citations in this group, jot down your thoughts about the group's common thread in the scratchpad. There is no need to erase what you've

already written, just append your latest thinking to your working notes. Now answer the *How cohesive is this group?*

- If you have been able to detect a common thread or central theme for this group, and ALL citations relate to that central theme, then select *a) All citations have something in common.*

- If you have been able to detect a common thread or central theme for this group, and MOST but not all citations relate to that central theme, then select *b) Most citations have something in common.*

- If just a few of the citations in the group have something in common but most citations do not, then select *c) A few citations have something in common.* You should also select c) if you detect multiple common threads, with a few citations belonging to each thread.

- If you cannot detect a common thread or central theme for the citations in this group, then select *d) There is no common thread among these citations.*

It may be difficult for you to decide between b) *Most citations have something in common* and c) *A few citations have something in common.* If you are on the fence, choose b).

**RULE OF THUMB**
**If you can't make up your mind between *b) Most* and *c) A few*, choose *b) Most*.**

### 7.2 Does this group represent a useful/appropriate way to cluster citations for this query?

Note: You will only answer this question if you have decided that the group is reasonably cohesive; i.e. you have selected a) or b) for the question *How cohesive is this group*.

Possible answers to the question *Does this group represent a useful/appropriate way to cluster citations for this query* are:

a) Yes, this group represents an appropriate cluster for this query.
b) No, this group does not represent an appropriate cluster for this query.

By useful/appropriate, we mean that the group's common theme represents a potentially useful or informative way to organize information for the user. A group doesn't need to be actually useful in its implementation, but it should reflect a reasonable, appropriate (as opposed to arbitrary) grouping of citations.

For instance, consider the query *What do people think about Pope Shenouda III?.* A group whose common thread is *Egyptian reactions to Pope Shenouda III* could be an

informative way to organize citations for this query. Similarly, a group whose common thread is *Western religious leaders' reactions to Pope Shenouda III* is potentially helpful.

On the other hand, a group whose common thread is *Citations containing 22 words* doesn't represent a potentially informative or useful grouping, because this common thread is arbitrary rather than appropriate to the query. Similarly, a group whose organizing theme is *Reactions to Pope Shenouda III by women whose names start with B* is not likely to be useful.

- If you believe that the group's common thread is a potentially useful and appropriate way to organize citations for this query, choose *Yes, this group represents an appropriate cluster for this query.*
- If the group's common thread seems arbitrary, not useful or otherwise inappropriate, choose *No, this group does not represent an appropriate cluster for this query*.

**RULE OF THUMB**
**If you can't make up your mind between *a) Yes* and *b) No*, choose *Yes, this group represents an appropriate cluster for this query*.**

### 7.3 Does the group label reasonably capture what these citations have in common?

Keeping in mind the common thread or central theme shared by the citations in this group, go back and look at the group label and answer the question *Does the group label reasonably capture what these citations have in common*?

Possible answers are

a) The group label is sufficiently informative and does not need to be changed.
b) The group label is somewhat informative but should be improved.
c) The group label is not informative and must be changed.

Examples of group labels include:

```
kind__gave__bush
merit__question__award
Citations related to: Bill Clinton's womanizing
Citations related to: Sarah Palin—LOL
Citations related to: Bob Cusack - *3-3-10 Mitt Romney
screens__concept__chinese
produces__price__simply
Citations related to: Palestinian Authority
Citations related to: joshphillips999 LOL
young__movies__violence
terrible__kill__makes
Citations related to: Love - Maria Mena –
```

As you can see from these examples, group labels are short, and you should think of them as titles or captions rather than verbose descriptions of the group's content. Group labels do not need to be complete grammatical sentences to be informative. Group labels can contain abbreviations and other kinds of shorthand, as long as you understand what is meant.

Note too that group labels may display particular formatting quirks, for instance using underscores between words rather than spaces and capitalization, or using a repetitive or formulaic phrase like "Citations related to: [topic]". You should not judge group labels harshly simply because you dislike the formatting or because structure is repeated from one label to the next. Instead, you should consider only the label's effectiveness in conveying information about the group's common thread.

It is not possible to judge group labels in isolation; labels can only be judged in the context of the common thread you have identified for this group. For instance, in the examples above, ***merit_question_award*** is a sufficiently informative label for a group of citations whose common thread is that they discuss the question of merit in Barack Obama's Nobel Peace Prize award. That same label would not be sufficiently informative for a group of citations whose common thread is that they discuss the health effects of war upon soldiers.

- If the group label serves as an adequate "title" this group's common thread, choose *The group label is sufficiently informative and does not need to be changed.* Choose this option for any group label that is sufficiently informative as to the group's common thread, even if you would prefer to see a slightly different formulation of the label.
- If the group label provides some, but not enough, information about the group's common thread, choose *The group label is somewhat informative but should be improved.*
- If the group label does not provide information about the group's common thread, choose *The group label is not informative and must be changed.*

**RULE OF THUMB**
**If you can't make up your mind between *a) sufficiently informative* and *b) somewhat informative*, choose *a) sufficiently informative*.**

**RULE OF THUMB**
**If you can't make up your mind between *b) somewhat informative* and *c) not informative*, choose *b) somewhat informative*.**

### 7.4 Provide an improved group label

If your answer to the question above is either

a) The group label is somewhat informative but should be improved.
b) The group label is not informative and must be changed.

you will be asked to provide an improved group label. Your improved group label should be a brief phrase, along the lines of a title or caption, that reflects the common thread shared by most/all citations in the group.

## 8. Citation-Level Assessment

Once you have answered all group-level questions you must judge each citation, using the citation panel in the middle of the annotation GUI. There is a single question to be answered for each citation:

**Is this citation an appropriate member of this group?**

Possible answers:

a) Yes.
b) No.

To answer this question, consider your final understanding of the group's common thread, which should be reflected in the (possibly improved) group label. For each citation, ask yourself whether it shares the group's common thread.

- If the citation shares the group's common thread, choose Yes.
- If the citation does not share the group's common thread, choose No.

**Important**: While answering questions about the group, you may have made preliminary judgments about the appropriateness of a citation's membership in the group. However, your understanding of the group and its theme or topic may have evolved during the course of assessment. It is very important that you confirm, for every citation, whether it belongs to the group given your final understanding of the group's theme after group assessment[2].

**RULE OF THUMB**
**If you can't make up your mind between *a) Yes* and *b) No*, choose *a) Yes*.**

## 9. View-Level Questions

Once you have completed assessment for all groups in this view (and all citations within each group), you need to answer several questions about the view as a whole. The annotation GUI will automatically present you with view-level questions when you have reached this stage.

[2] The annotation GUI allows assessors to record a preliminary yes/no "citation membership" judgment while answering group-level assessment questions. Citation-level judgments can be "finalized" only after all group-level assessments are completed, either by clicking the "final" checkbox next to an individual citation or by clicking the "finalize all citation judgments" button. The GUI permits assessors to sort citations by their status (yes vs. no vs. empty, final vs. non-final).

### 9.1 Does this view represent a useful/appropriate way to cluster groups for this query?

The first view-level question relates to the utility of the view. Possible answers to the question *Does this view represent a useful/appropriate way to cluster citations for this query* are:

a) Yes, this view represents an appropriate clustering of groups for this query.
b) No, this view does not represent an appropriate clustering of groups for this query.

In answering this question you should use the same principles that guided you in answering the corresponding group-level question.

While a group is a collection of citations, a view is a collection of groups. A useful/appropriate view will likely have some organizing principle for this collection of groups that is potentially useful or informative to the end user. For instance, a useful view might contain groups that each focus on a different person or geo-political entity. A useful view might contain groups representing different types of opinions about a topic. Keep in mind that a view doesn't need to be actually useful in its implementation, but it should reflect a reasonable, appropriate (as opposed to arbitrary) clustering of groups.

Recall that view labels are not required to be informative, so you should not base your answer on the view label. Instead, base your answer on the groups that comprise this view. Think about the common threads within each group, as reflected in the possibly improved group labels; then think about the groups taken as a whole and whether this collection of groups – the view – is a potentially useful way to present information to an end user.

- If this view (collection of groups) is a potentially useful way to organize information for this query then select *Yes, this view represents an appropriate clustering of groups for this query.*
- If this view (collection of groups) is arbitrary not a potentially useful way to organize information for this query then select *No, this view does not represent an appropriate clustering of groups for this query.*

**RULE OF THUMB**
**If you can't make up your mind between *a) appropriate* and *b) not appropriate*, choose *a) appropriate*.**

### 9.2 Does the view label reasonably capture what these groups have in common?

Keeping in mind the organizing principle for this view (i.e., this collection of groups), go back and look at the view label and answer the question *Does the view label reasonably capture what these groups have in common*?

Possible answers are

d) The view label is sufficiently informative and does not need to be changed.
e) The view label is somewhat informative but should be improved.
f) The view label is not informative and must be changed.

Examples of view labels include:

```
Groups displayed by: PERSON
Groups displayed by: GPE
x_m0.tf_1.euc0
Topics
x_m0.cai0.euc0
```

As with group labels, view labels are short, and you should think of them as titles or captions rather than verbose descriptions of the view's organizing principle. As with group labels, view labels do not need to be complete grammatical sentences and they may contain abbreviations or formatting or structural peculiarities. And finally, keep in mind that view labels can only be judged in the context of the organizing principle you have identified for this collection of groups.

- If the view label serves as an adequate "title" for the organizing principle for this collection of groups, choose *The view label is sufficiently informative and does not need to be changed.*
- If the view label provides some, but not enough, information about the organizing principle for this collection of groups, choose *The view label is somewhat informative but should be improved.*
- If the view label does not provide information about the organizing principle for groups in this view, choose *The view label is not informative and must be changed.*

**RULE OF THUMB**
**If you can't make up your mind between *a) sufficiently informative* and *b) somewhat informative*, choose *a) sufficiently informative*.**

**RULE OF THUMB**
**If you can't make up your mind between *b) somewhat informative* and *c) not informative*, choose *b) somewhat informative*.**

### 9.3 Provide an improved view label

If your answer to the question above is either

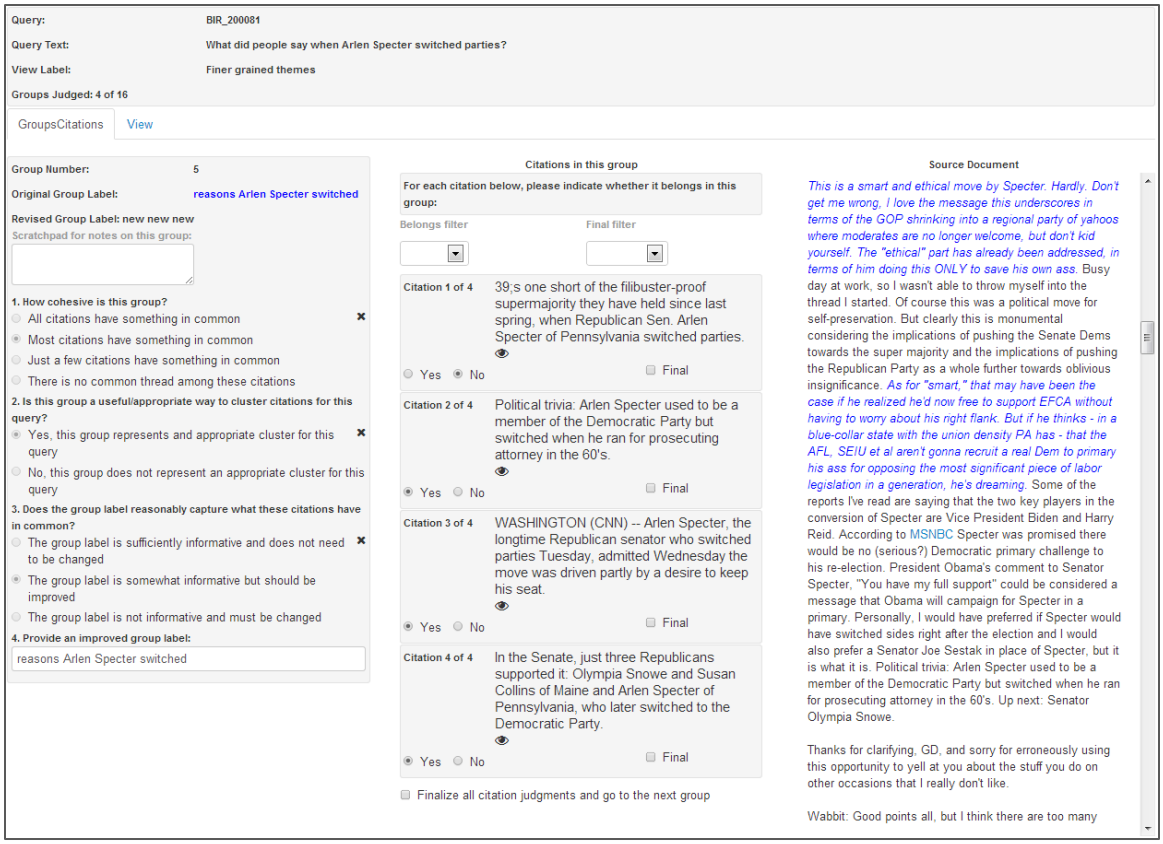
c) The view label is somewhat informative but should be improved.

d) The view label is not informative and must be changed.

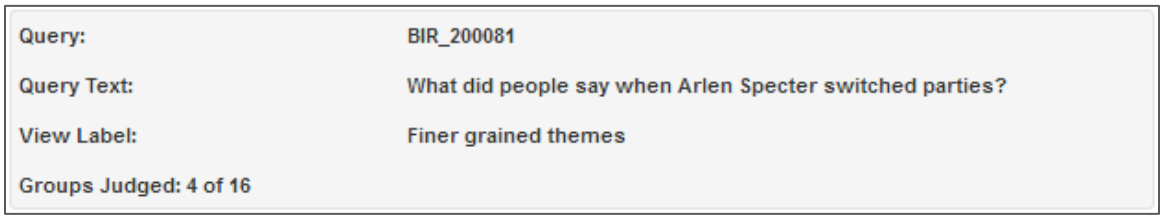
you will be asked to provide an improved view label. Your improved view label should be a brief phrase, along the lines of a title or caption, that reflects the organizing principle for this collection of groups.

## 10. The Annotation GUI

The following is a walkthrough of the annotation GUI in which thematic grouping assessment is performed. The GUI can be accessed by logging in at https://webann.ldc.upenn.edu/ using your webann username and password. Below is a full screenshot of the groups/citations tab, the primary assessment workspace.

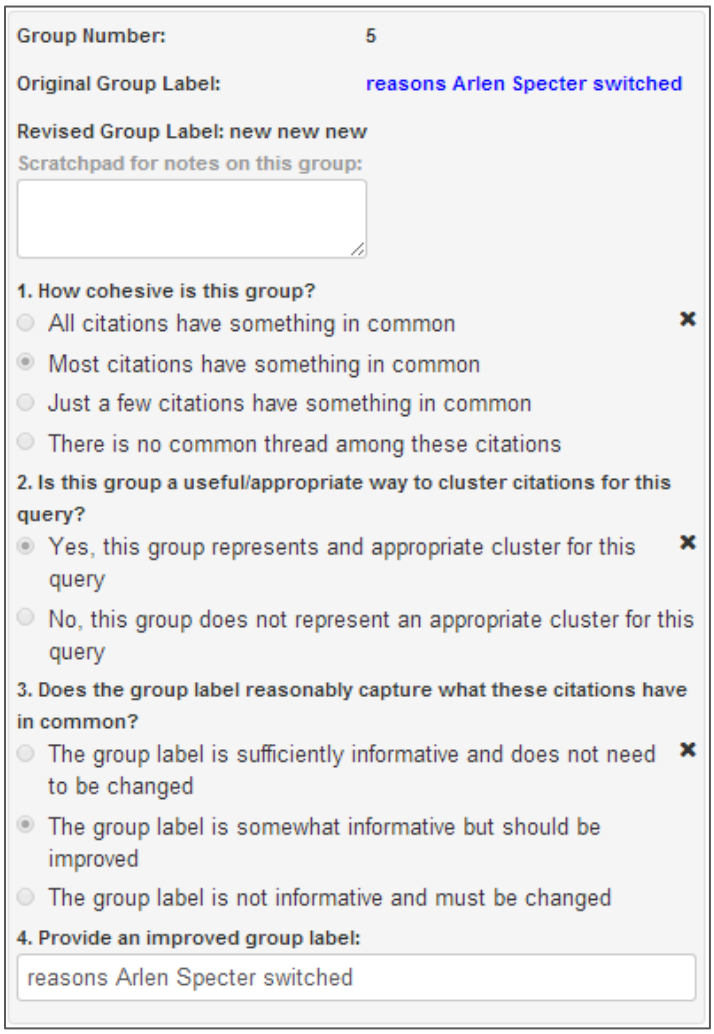



### 10.1 The Info Panel




The info panel, at the top of the GUI, informs the annotator of the query. Additionally, the overall progress within the view is displayed, as is the view label.

## 10.2 The Group Panel



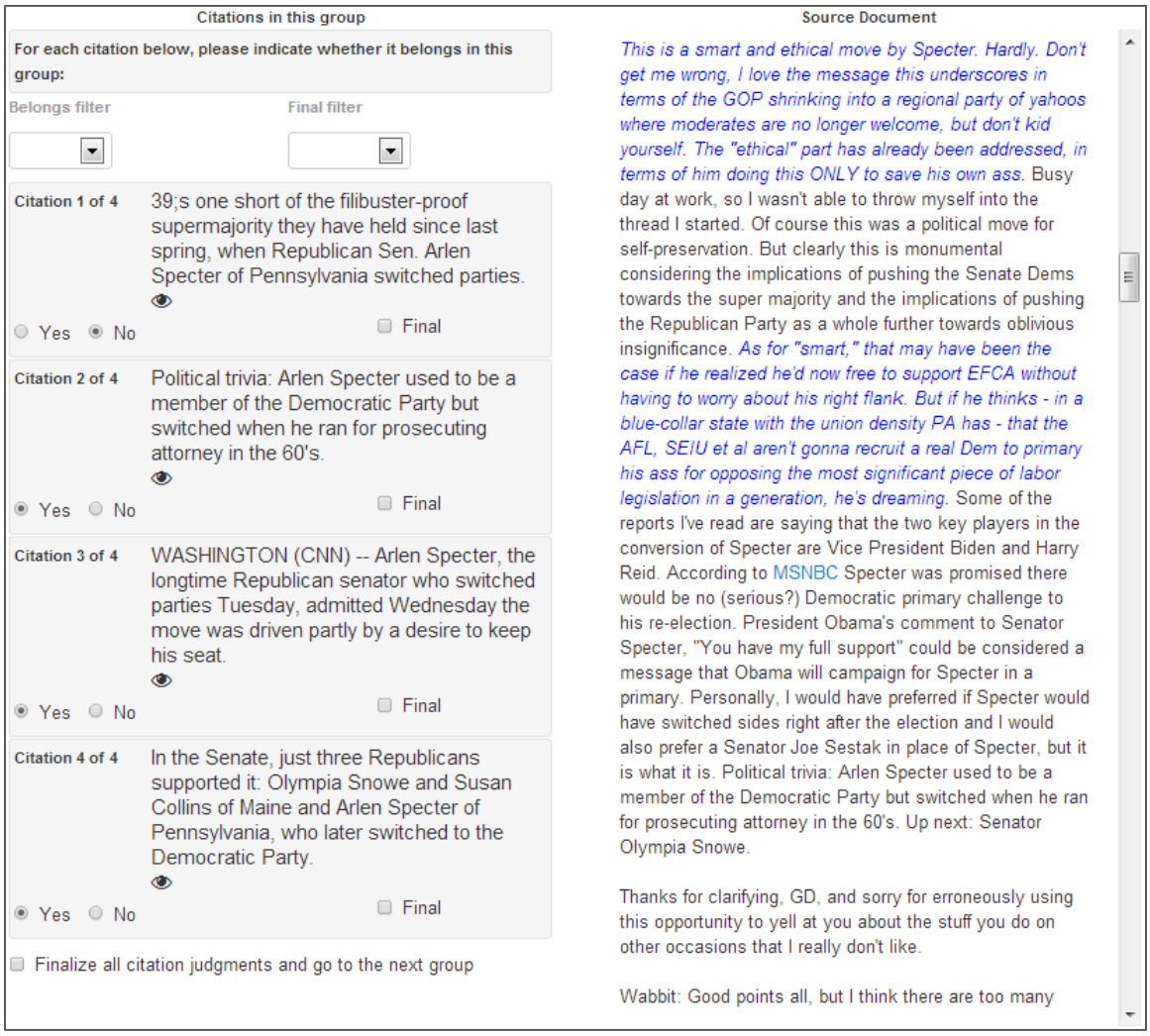

The group panel, the leftmost panel in the groups/citations tab, is where thematic grouping assessments are made. The GUI presents one question at a time. Whether a question requires an answer is dependent upon the answer selected for the preceding question. For instance, if for question 1, "There is no common thread among these citations" is selected, no further questions about the group will be displayed.

After a question has been answered, the GUI will disable the checkboxes beneath that question. However, note the ✖ buttons next to group questions 1, 2 and 3. After selecting an answer, one may click an ✖ button in order to reset that question (as well as any questions that follow), allowing the question to be revisited and a different answer selected.

Note also the scratchpad feature. Notes can be taken here regarding potential group themes, specific citations, etc. – whatever helps you organize your thoughts. Text entered in the scratchpad does not need to be deleted.

When question 3 is answered “somewhat informative” or “not informative”, question 4 will display, which requires a new group label to be entered. Simply click within the text box, type your improved group label, and hit enter. The group label at the top of the group panel will be updated to match (and will turn blue).

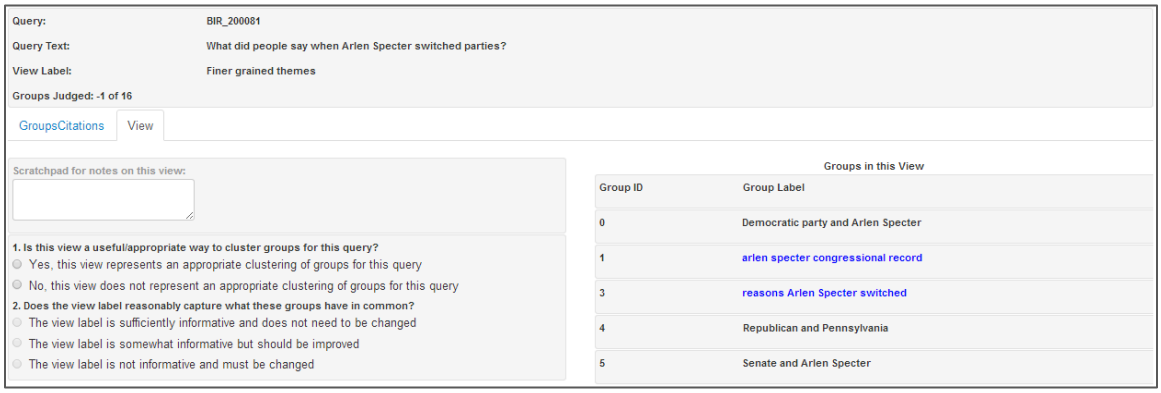
## 10.3 The Citation and Source Panels




The citation panel, the center panel in the groups/citations tab, is where decisions about individual citations are made. In order to finalize a decision, the “Final” checkbox must be checked. All decisions about citations must be finalized before moving onto the next group. Note the drop-down menus at the top of the citation panel. These menus will allow you to sort which citations are displayed. The “Belongs filter” menu filters citations by their assessment (“Yes” or “No”) and can also be used to show only those citations which have yet to be assessed (“Empty”). The “Final filter” menu filters citations by whether or not they’ve been finalized (“Final” or “Not Final”). Changing both menus back to no selection will always show the full list of citations.

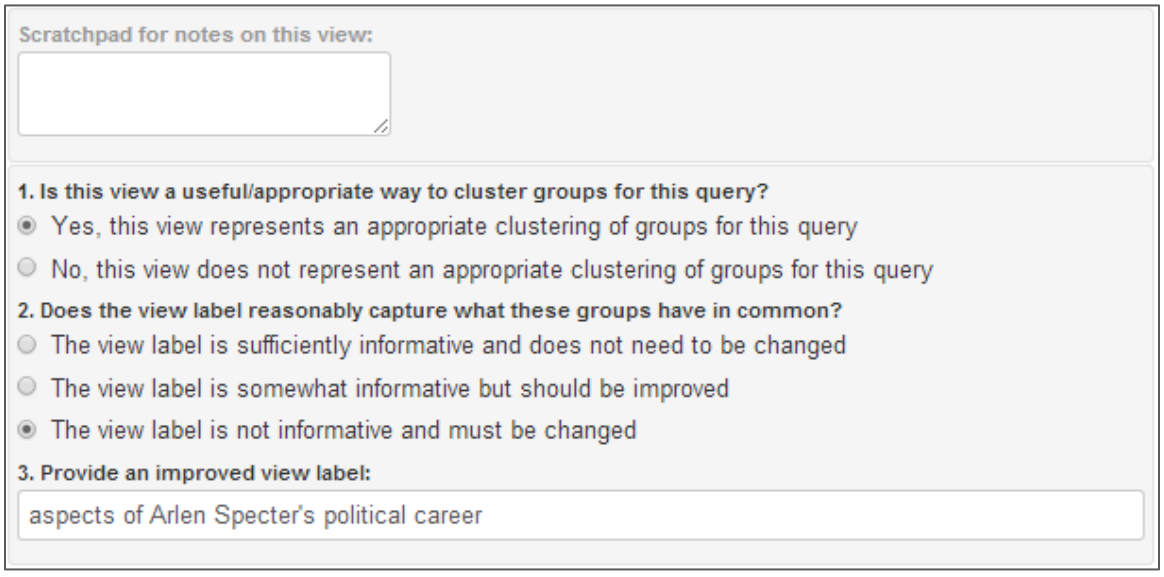
To view a citation in context, click on the button beneath the citation. This will cause the document from which the citation came to display in the source panel, the rightmost panel in the groups/citations tab. The citation text will be colored red in the source panel.

## 10.4 The View Tab




After all thematic grouping assessment is completed, **annotators must manually move to the view tab** to complete all view-level questions. The view tab is comprised of two panels. The right panel is where view-level questions are displayed and answered. The left panel displays a list of all groups within the view. Groups with improved group labels are in blue.

## 10.5 The View Questions Panel




Note that the view questions panel of the view tab also has a scratchpad in which notes can be kept regarding your thoughts about the theme of the view, etc. When question 2 is answered “somewhat informative” or “not informative”, question 3

will display, which requires a new view label to be entered. Simply click within the text box, type your improved view label, and hit enter.

When all view questions have been completed, click the “Next” button at the bottom of the page to mark the kit done and move on to a new kit.