

2018

## Automating CIRI Ratings of Human Rights Reports Using Gate

Joshua M. Joiner

University of North Florida, n00616193@ospreys.unf.edu

Follow this and additional works at: <https://digitalcommons.unf.edu/etd>



Part of the [Databases and Information Systems Commons](#), and the [Software Engineering Commons](#)

---

### Suggested Citation

Joiner, Joshua M., "Automating CIRI Ratings of Human Rights Reports Using Gate" (2018). *UNF Graduate Theses and Dissertations*. 784.

<https://digitalcommons.unf.edu/etd/784>

This Master's Thesis is brought to you for free and open access by the Student Scholarship at UNF Digital Commons. It has been accepted for inclusion in UNF Graduate Theses and Dissertations by an authorized administrator of UNF Digital Commons. For more information, please contact [Digital Projects](#).

© 2018 All Rights Reserved

AUTOMATING CIRI RATINGS OF HUMAN RIGHTS REPORTS  
USING GATE

by

Joshua Joiner

A thesis submitted to the  
School of Computing  
in partial fulfillment of the requirements for the degree of

Master of Science in Computing and Information Sciences

UNIVERSITY OF NORTH FLORIDA  
SCHOOL OF COMPUTING

May 2018

Copyright (©) 2018 by Joshua Joiner

All rights reserved. Reproduction in whole or in part in any form requires the prior written permission of Joshua Joiner or designated representative.

The thesis “Automating CIRI Ratings of Human Rights Reports Using Gate” submitted by Joshua Joiner in partial fulfillment of the requirements for the degree of Master of Science in Computing and Information Sciences has been

Approved by the thesis committee:

Date

---

Dr. Karthikeyan Umapathy  
Thesis Advisor and Committee Chairperson

---

Dr. Ching-Hua Chuan

---

Dr. Sandeep Reddivari

Accepted for the School of Computing:

---

Dr. Sherif Elfayoumy  
Director of the School

Accepted for the College of Computing, Engineering, and Construction:

---

Dr. Mark Tumeo  
Dean of the College

Accepted for the University:

---

Dr. John Kantner  
Dean of the Graduate School

## ACKNOWLEDGEMENT

I would like to give my thanks to Professor Karthikeyan Umapathy who guided and assisted me through this thesis process. I would also like to extend my thanks to committee members Professors Ching-Hua Chuan and Sandeep Reddivari for their advice and input in helping me improve this thesis.

## CONTENTS

List of Figures .....	vii
List of Tables .....	viii
List of Equations .....	x
Chapter 1. Introduction .....	1
Chapter 2. Background and Literature review .....	4
2.1    United States Human Rights Reports .....	4
2.2    Amnesty International.....	5
2.3    CIRI Human Rights Data Project .....	7
2.4    CIRI Dataset .....	8
2.5    Text Mining.....	14
2.6    Related Work.....	15
Chapter 3. Research method.....	18
3.1    Design Science Research Method .....	18
3.2    Design Science Research Guidelines .....	19
Chapter 4. GATE Text Mining platform .....	22
4.1    GATE Architecture .....	23
4.2    ANNIE: A Nearly-New Information Extraction System.....	24
4.3    Natural Language Understanding API .....	29
4.4    Annotation Diff Tool .....	29
4.5    GATE Embedded.....	36

Chapter 5. Implementation .....	37
5.1 Step One – Parsing, Extracting, and Annotating.....	39
5.2 Step Two – Executing Custom Processing Resources.....	40
5.3 Step Three – Generating and Storing CIRI Ratings.....	44
Chapter 6. Evaluation .....	46
6.1 Step One – Store the Source and Automated CIRI Data into Tables.....	48
6.2 Step Two – Retrieve CIRI Data to Create Annotation Sets.....	48
6.3 Step Three – Evaluate the Annotation Sets .....	49
6.4 Data Collection.....	50
6.5 Trends and Analysis.....	51
6.5.1 Four Human Rights Practice Groups .....	53
6.5.2 Independent Judiciary .....	58
6.5.3 Physical Integrity .....	60
6.5.4 Empowerment Rights .....	64
6.5.5 Women's Rights .....	71
6.5.6 Results Summary .....	73
Chapter 7. Conclusion .....	75
References.....	78
Appendix A.....	81
Appendix B.....	83
Appendix C .....	88
Vita.....	95

## FIGURES

Figure 1.	U.S. DoS Country Report Structure .....	4
Figure 2.	U.S DoS Country Report Content .....	5
Figure 3.	Amnesty International Country Report Structure .....	6
Figure 4.	Amnesty International Country Report Content .....	6
Figure 5.	CIRI Dataset Structure .....	9
Figure 6.	GATE Architecture .....	24
Figure 7.	Standard ANNIE process flow .....	25
Figure 8.	Custom ANNIE process flow .....	26
Figure 9.	Plain text document.....	27
Figure 10.	Text tagged with Annotations.....	28
Figure 11.	Checked Annotations .....	28
Figure 12.	Annotation List of Annotation metadata.....	29
Figure 13.	Standard Precision/Recall Diagram.....	32
Figure 14.	GATE Precision/Recall Diagram .....	33
Figure 15.	Automated Development Process .....	37
Figure 16.	GATE Application PR Structure .....	39
Figure 17.	CIRI Database Table Structure.....	44
Figure 18.	Evaluation Plan .....	47
Figure 19.	CIRI Document Comparison.....	50



## TABLES

Table 1.	Human Rights Practices Indicators .....	10
Table 2.	Human Practice Ratings .....	11
Table 3.	Identity Indicators .....	12
Table 4.	Human Rights Practice Definitions .....	13
Table 5.	Precision/Recall Example.....	35
Table 6.	Evaluation Sample Countries .....	53
Table 7.	Denmark Independent Judiciary Results.....	55
Table 8.	Denmark Physical Integrity Results .....	56
Table 9.	Denmark Empowerment Rights Results .....	56
Table 10.	Denmark Women’s Rights Results .....	57
Table 11.	Denmark Overall Evaluation Results .....	57
Table 12.	Canada Independent Judiciary Key and Response Data.....	58
Table 13.	Cambodia Independent Judiciary Key and Response Data .....	59
Table 14.	Bahrain Independent Judiciary Key and Response Data .....	59
Table 15.	Angola Physical Integrity Key and Response Data .....	61
Table 16.	Angola Physical Integrity F-Measure Scores.....	61
Table 17.	Sudan Physical Integrity Key and Response Data .....	62
Table 18.	Sudan Physical Integrity F-Measure Scores .....	62
Table 19.	Colombia Physical Integrity Key and Response Data .....	63
Table 20.	Colombia Physical Integrity F-Measure Scores .....	63
Table 21.	Austria Empowerment Rights Key and Response Data .....	65

Table 22.	Austria Empowerment Rights F-Measure Scores.....	66
Table 23.	Austria Overall Evaluation Results .....	66
Table 24.	Bhutan Empowerment Rights Key and Response Data .....	67
Table 25.	Bhutan Empowerment Rights F-Measure Scores .....	68
Table 26.	Bhutan Overall Evaluation Results .....	68
Table 27.	Pakistan Empowerment Rights Key and Response Data .....	69
Table 28.	Pakistan Empowerment Rights F-Measure Scores .....	70
Table 29.	Pakistan Overall Evaluation Results .....	70
Table 30.	Australia Women’s Rights Key and Response Data .....	71
Table 31.	Australia Women’s Rights F-Measure Scores .....	72
Table 32.	Egypt Women’s Rights Key and Response Data.....	72
Table 33.	Egypt Women’s Rights F-Measure Scores .....	73

## EQUATIONS

Equation 1.	Standard Precision Formula .....	31
Equation 2.	Standard Recall Formula .....	31
Equation 3.	GATE Precision Formula .....	33
Equation 4.	GATE Recall Formula.....	33
Equation 5.	F-measure.....	34

## ABSTRACT

This thesis involves parsing document-based reports from the United States Human Rights Reports and rating the human practices for various countries based on the CIRI (Cingranelli-Richards) Human Rights Data Project dataset. The United States Human Rights Reports are annual reports that cover internationally recognized human rights practices regarding individual, civil, political, and worker rights. Students, scholars, policymakers, and analysts used the CIRI data for practical and research purposes. CIRI analyzed the annual reports from 1981 to 2011 and then stopped releasing the dataset for any further years, but a possible reason is due to the manual process of scouring the Human Rights Reports and then rating each human rights practice for each country. This manual process provides a solid foundation for creating a new automated process. The automated process uses the rating values provided by CIRI in the 1981-2011 dataset as expected values to evaluate the accuracy of the rating process.

To transition to an automated process, the General Architecture for Text Engineering (GATE) application is used. GATE is an open source project used for developing solutions for text processing. GATE is used in conjunction with the coding schemes provided within the CIRI Coding Manual to create an automated ratings process. The CIRI Coding Manual uses qualitative and quantitative criteria. The original and automated ratings are evaluated using GATE's Annotation Diff Tool to get the

F-measure for every country in the dataset. The evaluation cases range between 1999 and 2011 because those are the only years included in both the CIRI dataset and the Human Rights Reports. The F-measure results are more accurate when quantitative criteria is used to rate human rights practices. The primary contribution of this thesis is a method for automating each country's human practice ratings so that the purpose of the CIRI project can be continued.

## Chapter 1

### INTRODUCTION

The United States Department of State (U.S. DoS) releases annual country reports on human rights practices called the Human Rights Reports to help promote and protect universal human rights and support efforts to hold governments accountable. These reports, first issued in 1977, cover internationally recognized rights that are set by the Universal Declaration of Human Rights and international agreements, which include individual, civil, political, and worker rights. Currently, the United States Department of State website only has reports available for years 1999 to 2016. The United States Department of State submits these annual reports to the United States Congress on all countries that are receiving assistance from U.S., as well as all United Nation member states [United States Department of State<sup>16</sup>].

Another relevant organization, Amnesty International that was founded in 1961 also releases annual country reports on human rights violations for over 100 countries, but it is a non-governmental organization and is supported by more than 7 million people globally. The organization researches and exposes facts detailing international human rights violations perpetrated by both corporations and governments [Amnesty International<sup>16</sup>]. These reports help activists, governmental agencies, and special interest groups keep their promises and respect of international law.

The CIRI Human Rights Data Project [Cingranelli14A] was a project that annually rated the level of government respect for 204 countries based on various internationally recognized human rights. The project is now discontinued, but from 2004 to 2014 the project published ratings of these countries in a dataset that is still available on their website at <http://www.humanrightsdata.com/p/data-documentation.html>. The website is where the current and final dataset that includes data from years ranging from 1981 to 2011 can be downloaded. The project specifically measured the human rights practices and not the government policies or human rights conditions. The project was created to be used by scholars, students, and analysts for academic and research purposes regarding causes and consequences of human rights violations. The CIRI Human Rights Data Project used the country information contained from the Human Rights Reports from the United States Department of State and Amnesty International to produce the country ratings [Cingranelli14A].

The groups using CIRI's dataset to analyze human rights effects on various institutions no longer have access to updated country human rights practice ratings since the discontinuation of the CIRI Human Rights Data Project. CIRI's rating process involved manual rating of the human rights practices by their staff members. However, CIRI has posted their rating schemes on their website. This provides an opportunity to automate CIRI's rating process, which is the research objective for this thesis.

In this thesis, the natural language processing software GATE is used to automate CIRI's manual process. Using the coding schemes provided in CIRI's documentation, patterns

are created to extract the necessary information from the Human Rights reports using the tools provided by GATE. The goal is to develop an automated system that provides accurate ratings that are similar to the ratings in the CIRI dataset for years 1999-2011. The utility of the automated tool will be evaluated by comparing its results against the CIRI dataset to determine the accuracy of the ratings. If evaluation results reveal that the automated tool is highly accurate, then the results can be used as basis for future analysis of human practice ratings.



## Chapter 2

### BACKGROUND AND LITERATURE REVIEW

#### 2.1 United States Human Rights Reports

Each annual country report that the United States Department of State releases contains several sections of information pertaining to specific human rights practices. The country reports are structured with an Executive Summary and seven sections which may include subsections [United States Department of State16]. Figure 1 below provides a complete listing of the sections and sub-sections within a typical country report [United States Department of State16].

Select the sections to include:

<input type="checkbox"/> Executive Summary	<input type="checkbox"/> Section 2. Respect for Civil Liberties, Including:	<input type="checkbox"/> Section 6. Discrimination, Societal Abuses, and Trafficking in Persons	<input type="checkbox"/> Section 7. Worker Rights
<input type="checkbox"/> Section 1. Respect for the Integrity of the Person, Including Freedom from:	<input type="checkbox"/> a. Freedom of Speech and Press	<input type="checkbox"/> Women	<input type="checkbox"/> a. Freedom of Association and the Right to Collective Bargaining
<input type="checkbox"/> a. Arbitrary or Unlawful Deprivation of Life	<input type="checkbox"/> b. Freedom of Peaceful Assembly and Association	<input type="checkbox"/> Children	<input type="checkbox"/> b. Prohibition of Forced or Compulsory Labor
<input type="checkbox"/> b. Disappearance	<input type="checkbox"/> c. Freedom of Religion	<input type="checkbox"/> Anti-Semitism	<input type="checkbox"/> c. Prohibition of Child Labor and Minimum Age for Employment
<input type="checkbox"/> c. Torture and Other Cruel, Inhuman, or Degrading Treatment or Punishment	<input type="checkbox"/> d. Freedom of Movement, Internally Displaced Persons, Protection of Refugees, and Stateless Persons	<input type="checkbox"/> Trafficking in Persons	<input type="checkbox"/> d. Discrimination with Respect to Employment or Occupation
<input type="checkbox"/> Prison and Detention Center Conditions	<input type="checkbox"/> Section 3. Freedom to Participate in the Political Process	<input type="checkbox"/> Persons with Disabilities	<input type="checkbox"/> e. Acceptable Conditions of Work
<input type="checkbox"/> d. Arbitrary Arrest or Detention	<input type="checkbox"/> Elections and Political Participation	<input type="checkbox"/> National/Racial/Ethnic Minorities	
<input type="checkbox"/> e. Denial of Fair Public Trial	<input type="checkbox"/> Section 4. Official Corruption and Government Transparency	<input type="checkbox"/> Indigenous People	
<input type="checkbox"/> f. Arbitrary Interference with Privacy, Family, Home, or Correspondence	<input type="checkbox"/> Section 5. Governmental Attitude Regarding International and Nongovernmental Investigation of Alleged Violations of Human Rights	<input type="checkbox"/> Acts of Violence, Discrimination, and Other Abuses Based on Sexual Orientation and Gender Identity	
<input type="checkbox"/> g. Use of Excessive Force and Other Abuses in Internal Conflicts		<input type="checkbox"/> HIV and AIDS Social Stigma	
		<input type="checkbox"/> Other Societal Violence or Discrimination	
		<input type="checkbox"/> Promotion of Acts of Discrimination	

Figure 1. U.S. DoS Country Report Structure

Figure 2, as shown below, provides a snippet of the type of content contained within the country report sections [United States Department of State16]. The example below provides a portion of text from a sub-section in section 1 of the 2015 Afghanistan report. This text explains the human rights violations occurring in Afghanistan committed by government or entities related to government.



Figure 2. U.S. DoS Country Report Content

## 2.2 Amnesty International

The independence of the Amnesty International organization provides another source for country reports and information on human rights practices. In Amnesty International reports, countries are grouped by region, so that each region receives an overview section, as shown in Figure 3 [Amnesty International16]. After the regional overview section, there are detailed reports on each country, as shown in Figure 4 [Amnesty International16].

<b>CONTENTS</b>	
<b>ANNUAL REPORT 2015/16</b>	
ABBREVIATIONS .....	ix
PREFACE .....	xi
FOREWORD .....	14
AFRICA REGIONAL OVERVIEW .....	18
AMERICAS REGIONAL OVERVIEW .....	26
ASIA-PACIFIC REGIONAL OVERVIEW .....	34
EUROPE AND CENTRAL ASIA REGIONAL OVERVIEW .....	41
MIDDLE EAST AND NORTH AFRICA REGIONAL OVERVIEW .....	49
AFGHANISTAN .....	60
ALBANIA .....	63
ALGERIA .....	65

Figure 3. Amnesty International Country Report Structure

Figure 4, as depicted below shows the individual country section and the accounts of human rights violations as measured by Amnesty International [Amnesty International16]. It is important to note that the information reported by both the United States Department of State and Amnesty International will vary because each organization acts independently.

<b>AFGHANISTAN</b>	
<b>Islamic Republic of Afghanistan</b> Head of state and government: <b>Muhammad Ashraf Ghani Ahmadzai</b>	Security. It enshrined the government's pledges to increase women's role in the four pillars of Resolution 1325: participation, protection, prevention, and relief and recovery.
<b>There was growing insecurity with insurgency and criminal activity worsening across the country. The first three months of 2015 were the most violent of any equivalent period on record. The UN Assistance Mission in Afghanistan (UNAMA) recorded 1,592 civilians killed and 3,329 injured in the first six months of 2015, while 70% of civilian casualties were attributed to Taliban and other armed insurgent groups, and 16% to pro-Afghan government forces. The Taliban increasingly</b>	On 29 July the government proclaimed that Mullah Omar, the Taliban leader, died in April 2013 in Pakistan. Following this announcement a string of attacks occurred in the capital, Kabul, between 7 and 10 August. Mullah Akhtar Mohammad Mansoor, Mullah Omar's deputy since 2010, was announced as his successor. In his first public statement as the new leader on 1 August, he called for Taliban unity and continued jihad, while characterizing reports of a peace process as enemy propaganda. In May, the Ministry of Interior estimated that there were some 7,180 foreign fighters across Afghanistan, the

Figure 4. Amnesty International Country Report Content

### 2.3 CIRI Human Rights Data Project

The CIRI Human Rights Data Project provided annual updates about government respect for a variety of human rights practices for over 200 countries. The founders of the project, Dr. David Cingranelli of Binghamton University and Dr. David L. Richards of the University of Connecticut, were political scientists. In 2013, Dr. K. Chad Clay of the University of Georgia joined the project as a co-director. One of the main purposes of the project was “to expand theory building and empirical research about government human rights practices beyond the extant dominant concern with violations of physical integrity rights” [Cingranelli14A]. The author’s believed that by measuring the government respect of human rights practices we can begin to understand the causes and effects of human rights violations on “institutional changes and public policies including democratization, economic aid, military aid, structural adjustment, and humanitarian intervention” [Cingranelli14A]. In 1994, the project was originally created for those who study government human rights practices, but then became more widely used by governments and other organizations, including intergovernmental organizations, non-governmental, think-tanks, and businesses for domestic and international policymaking [Cingranelli14A]. It should be noted that the CIRI Human Rights Data Project was discontinued in 2014.

The indicators used in the CIRI dataset were created using a mixed-methods approach by using content analysis of qualitative material that describes respect for human rights in each country. The coding criteria was developed in a way to reflect both the meanings of various human rights as defined by the international human rights law and to represent the

various ways that human rights law and government behavior actually intersect. The human rights ratings used in the CIRI project are considered standards-based because the scores reflect the ratings of actual government practices relative to the standards set by international law, particularly the Universal Declaration of Human Rights. To produce standardized ratings of human rights practices, CIRI coders use the annual reports provided by the United States Department of State's Country Reports on Human Rights Practices. The paper states that the ratings given to each human rights practice is ordinal in measurement because human rights information is imperfect [Cingranelli14A]. The CIRI coders use the annual reports of Amnesty International as a second authoritative source for human rights regarding extrajudicial killing, disappearance, torture, and political imprisonment [Cingranelli14B].

CIRI coding experts and senior staff read the human rights reports to create CIRI data points. At least two coders were tasked to code one CIRI data point based on the specific coding schemes for that data point [Cingranelli14A]. For example, a human rights practice indicator may be labeled KILL for the number of governmental related killings that occurred in Canada in 2011. When all of the data points were completed for a given year, the complete CIRI dataset was produced.

## 2.4 CIRI Dataset

The CIRI dataset defined short variable descriptions for 27 indicators, several of which are retired. Coding experts read the annual country reports and then give initial ratings, which

are then reviewed by CIRI senior staff. Figure 5, a sample taken from the CIRI dataset, shows an example of how the dataset is structured for storing the rating values for each country [Cingranelli14C].

1	CTRY	YEAR	CIRI	COW	POLITY	UNCTRY	UNREG	UNSUBREG	PHYSINT	DISAP	KILL	POLPRIS	TORT
2	Afghanistan	1981	101	700	700	4	142	62	0	0	0	0	0
3	Afghanistan	1982	101	700	700	4	142	62	0	0	0	0	0
4	Afghanistan	1983	101	700	700	4	142	62	0	0	0	0	0
5	Afghanistan	1984	101	700	700	4	142	62	0	0	0	0	0
6	Afghanistan	1985	101	700	700	4	142	62	0	0	0	0	0
7	Afghanistan	1986	101	700	700	4	142	62	0	0	0	0	0

Figure 5. CIRI Dataset Structure

Most of the ratings for human rights practices, which are listed below in Table 1, range from 0 to 2 and are shown in Table 2 along with what each rating represents. There are, however, exceptions regarding several categories of women's rights where the rating ranges from 0 to 3, also shown in Table 2, where 0 is no women's rights by law and 3 is equal women's rights by law and in practice. The other exceptions are cumulative identifiers that are the sum of one or more human rights practice ratings [Cingranelli14B].

The Empowerment Rights Index has a rating between 0 and 12 and the Physical Integrity Rights Index has a rating between 0 and 8. Missing data has a "-999" code, the "-77" code indicates periods of interregnum, and code "-66" indicates periods of interruption. More information about the indicators can be found in the document for the CIRI Short Variable Descriptions on their website [Cingranelli14A]. The Independent Judiciary and Women's Rights are custom classifications that were created for this thesis research purposes for aiding evaluation and presentation of results; and these custom classifications do not affect

the CIRI coding scheme found in the CIRI Human Rights Data Project Coding Manual in any way [Cingranelli14B].

Indicator	Human Rights Practice
<b>Practice Indicators</b>	
ASSN	Freedom of Assembly & Association
DISAP	Disappearance
DOMMOV	Freedom of Domestic Movement
ELECS	Electoral Self-Determination
FORMOV	Freedom of Foreign Movement
INJUD	Independence of the Judiciary
POLPRIS	Political Imprisonment
KILL	Political or Extrajudicial Killing
NEW_RELFRE	Freedom of Religion
SPEECH	Freedom of Speech
TORT	Torture
WECON	Women's Economic Rights
WOPOL	Women's Political Rights
WORKER	Workers' Rights
<b>Index Indicators</b>	
NEW_EMPINX	Empowerment Rights Index (Cumulative Index for: electoral self-determination, domestic movement, foreign movement, religion, speech, assembly & association, and workers' rights)
PHYSINT	Physical Integrity Rights of Index (Cumulative Index for: disappearance, extrajudicial killing, political imprisonment, and torture)
<b>Custom Groups</b>	
INDEPENDENT JUDICIARY	Independence of the Judiciary
WOMEN'S RIGHTS	Women's Economic Rights, Women's Political Rights

Table 1. Human Rights Practices Indicators

Rating	Representation
<b>Common Human Practice Ratings</b>	
0	Frequent violations of this right
1	Some violations of this right
2	No reported violations of this right
<b>Women's Rights Human Practice Ratings</b>	
0	No rights granted by law. Government tolerates high level of discrimination against women.
1	Some rights granted. Government does not enforce laws effectively in practice. Government tolerates moderate level of discrimination against women.
<b>Women's Rights Human Practice Ratings</b>	
2	Some rights granted. Government enforces laws effectively in practice. Government tolerates low level discrimination against women.
3	All or nearly all of rights guaranteed by law. Government fully enforces laws in practice. Government tolerates none or almost no discrimination against women.

Table 2. Human Practice Ratings

As seen in the sample dataset above in Figure 5, each country has a set of indicators that are listed before the human practices ratings are listed. The first eight columns in the dataset are considered as identifiers, which include the country, year, a custom CIRI identifier, and various codes. See Table 3 for listing of dataset identifiers [Cingranelli14A].



<b>INDICATOR</b>	<b>IDENTIFIER NAME</b>
CTRY	COUNTRY NAME
YEAR	YEAR IDENTIFIER
CIRI	CIRI COUNTRY IDENTIFIER
COW	CORRELATED OF WAR
POLITY	POLITY COUNTRY IDENTIFIER
UNCTRY	UNITED NATIONS COUNTRY IDENTIFIER
UNREG	UNITED NATIONS REGION IDENTIFIER
UNSUBREG	UNITED NATIONS SUB REGION IDENTIFIER

Table 3. Identity Indicators

Coders use the CIRI Coding Manual located on their website to determine ratings for the human rights assigned to them [Cingranelli14B]. The guide provides more information on the coding schemes and the sections from the United State Department of State Country Reports needed to obtain information and other dos and don'ts [Cingranelli14B]. Table 4 lists the relations between the human rights practices rated by CIRI and the corresponding section(s) in the U.S. human rights country report [Cingranelli14B].

<b>CIRI Human Rights Practice</b>	<b>U.S. Human Rights Country Report Section</b>
Political or Extrajudicial Killing	Section One (Respect for the Physical Integrity of the Person, Including Freedom From:), Subsection A (Arbitrary or Unlawful Deprivation of Life / Political and Other Extrajudicial Killing).
Disappearance	Section One (Respect for the Physical Integrity of the Person, Including Freedom From:), Subsection B (Disappearance).
Torture	Section One (Respect for the Physical Integrity of the Person, Including Freedom From:), Subsection C (Torture and Other Cruel, Inhuman, or Degrading Treatment or Punishment) and Subsection D (Arbitrary Arrest or Detention: Arrest Procedures and Treatment While in Detention).
Political Imprisonment	Section One (Respect for the Physical Integrity of the Person, Including Freedom From:), Subsections D (Arbitrary Arrest or Detention), and E (Denial of Fair Public Trial); and Section Two (Respect for Civil Liberties, Including:), Subsection A (Freedom of Speech and Press).
Freedom of Speech and Press	Section Two (Respect for Civil Liberties, Including:), Subsection A (Freedom of Speech and Press).
Freedom of Religion	Section Two (Respect for Civil Liberties, Including:), Subsection C (Freedom of Religion).
Freedom of Domestic Movement	Section Two (Respect for Civil Liberties, Including:), Subsection D: (Freedom of Movement, Internally Displaced Persons, Protection of Refugees, and Stateless Persons). Sometimes information is also included in Section 6 (Discrimination, Societal Abuses, and Trafficking in Persons) under "Women."
Freedom of Foreign Movement and Travel	Section Two (Respect for Civil Liberties, Including:), Subsection D: (Freedom of Movement, Internally Displaced Persons, Protection of Refugees, and Stateless Persons). Sometimes information is also included in Section 6 (Discrimination, Societal Abuses, and Trafficking in Persons) under "Women."
Freedom of Assembly and Association	Section Two (Respect for Civil Liberties), Subsection B (Freedom of Peaceful Assembly and Association).
Electoral Self-Determination	Section 3 (Respect for Political Rights: The Right of Citizens to Change Their Government).
Worker Rights	Section 2 (1981-1985), Section 5 (1986-1987), or Section 6 (1988-Present; Worker Rights).
Women's Political Rights	Section 3 (Respect for Political Rights: The Right of Citizens to Change Their Government).
Women's Economic Rights	Section 5 (Discrimination Based on Race, Sex, Religion, Disability, Language, or Social Status) and Section 6 (Worker's Rights). Sometimes, there is a "Women" subheading in Section 5.
Women's Social Rights (Discontinued as of 2005/2007)	Section 5 (Discrimination Based on Race, Sex, Religion, Disability, Language, or Social Status).
Independent Judiciary	Section One (Respect for the Physical Integrity of the Person, Including Freedom From:), Subsection E (Denial of Fair Public Trial).

Table 4. Human Rights Practice Definitions

## 2.5 Text Mining

Text mining is the process of extracting information from text and providing value [Hearst13]. Text mining involves parsing through the input, creating patterns to extract information from the input, and then evaluating the extracted information. Text mining techniques include text categorization, concept and entity extraction, and sentiment analysis.

Text categorization or document classification is a technique used to assign text to predefined categories. This technique is used to provide conceptual views on the subject(s) that are assigned to the category [Yang08]. For example, text about American football players may be categorized by the team they currently play for. Concept extraction or concept mining is different from text categorization because the process instead analyzes text and converts the words found within the text into concepts [Nilesh09]. These concepts can be types of entities, events, or topics. Entity extraction techniques are used to search through the text to find words and assign them into the appropriate category in the predefined list of persons, locations, numbers, etc. [Techopedia16]. Sentiment analysis, also called opinion mining, is used to analyze people's opinions, sentiments, and emotions towards products, events, or companies [Liu12].

Text mining applications help with the process of textual data. Some text mining techniques involve lexical analysis and annotating. Lexical analysis involves parsing a string of characters into tokens. A token is a string of characters that has some meaning

attached to it. For example, the word “reads” becomes a token that has a part-of-speech meaning attached to it, which would include metadata that denotes the string as an action verb with the root word of “read” [Farrell16]. This metadata can also be annotated and reference specific parts of the source text. See section 4.2 for an example of how GATE annotates text. GATE is used for its Natural Language Processing capabilities to extract information from the human rights reports provided by the United States of Department of State and Amnesty International.

## 2.6 Related Work

Minhas, Ulfelder, and Ward [Minhas15] demonstrated that using text mining and machine learning techniques on the United States Department of State Human Rights Country Reports can be considered as an effective means of data extraction. They explained that by using existing data and text they can train classifiers to predict the regime type of a country. Another paper by Rod Alence titled Mining for Meaning [Alence15] delved into the country reports produced by the African Peer Review Mechanism by using the “bag of words” text mining technique to find word frequency and distribution throughout the reports that give insight into themes of the reports and the concepts that are emphasized.

GATE, in particular, is used in various applications that need a robust set of text mining techniques. Diana Maynard and Mark Greenwood, in partnership with The UK National Archives, a non-ministerial government department, created a system that used GATE and

linked open data tools from Ontotext to help people access government website records [Maynard12]. The system provided annotations for common entities such as location, people, dates, and more specific annotations such as government departments, politicians, and civil servants. The methodology comprised of annotating entities within the text by using ontology-based information extraction tools to generate annotations, using GATE Mimir to search and index documents based on the generated annotations, and SPARQL to query for results. Their semantic annotation component achieved about 83% F-measure score [Maynard12].

Similar projects include using GATE to extract named entities from the XML tag descriptor of Yahoo RSS documents and store them as social networks [Mekala08] and to enhance the browsing experience of a digital library by implementing more ways information can be located and presented [Witten04]. Both of these projects use information extraction and annotation components provided by GATE's flexible architecture, discussed in Chapter 4. These projects show that not only is text mining important, but text mining applications with customizable features encourage the development of varying types of projects.

Research into this subject has revealed that even though there is a lot of studies regarding text mining or text analysis, there has been little regards to how text mining can be used in extracting data from reports published by various organizations on human rights practices. These findings provided an opportunity to fill the gap in the published literature and use the CIRI dataset in conjunction with the country reports and apply text mining techniques to

automate accurate replication of rating country's human rights practices. The next chapter discusses the research method, design science used in this thesis.

## Chapter 3

### RESEARCH METHOD

The type of research method used in this thesis is Design Science. The following sections explain what the Design Science research method is and the guidelines for creating a thesis based on this methodology.

#### 3.1 Design Science Research Method

Design science involves the creation and evaluation of artifacts that are intended to solve problems in Information Systems by improving performance or developing an understanding of the system. Some design science artifacts include algorithms, software, or natural language descriptions.

Design science research in Information Systems addresses the following problems stated by Hevner, March, Park, and Ram [Hevner04]:

- Unstable requirements and constraints being built upon ill-defined environmental contexts,
- Complex set of interactions between the subcomponents of the problem and solution,
- Design processes and artifacts that require flexibility to change,
- Solutions that require creative solutions, and

- Dependence on teamwork to produce effective solutions.

### 3.2 Design Science Research Guidelines

Hevner et al. establish the following guidelines to assist the community in understanding the requirements for effective design science research [Hevner04].

#### **Guideline 1 – Design as an Artifact**

The first guideline states that the design must produce a viable artifact in the form of a construct, a model, a method, or an instantiation. The focus of this thesis is to create a system for automating CIRI human practice rating values. The details of this artifact are discussed in the Implementation and Evaluation chapters. Thus, this thesis follows the design as an artifact guideline by designing an automated process for CIRI human practice rating values where the process was previously done manually.

#### **Guideline 2 – Problem Relevance**

The second guideline states that the objective of a design science research should be to develop technology-based solutions to important and relevant business problems. The automated rating value process can reduce resources, such as money and time, when parsing the country reports and evaluating the text to provide rating values. The relevance of this problem is established in the Introduction chapter.



### **Guideline 3 – Design Evaluation**

The third guideline states that the utility, quality, and efficacy of a design artifact must be rigorously demonstrated via well-executed evaluation methods. This thesis follows the design evaluation guideline as human rights practice ratings generated by the CIRI Automated System was evaluated using F-measure scores with CIRI dataset as the gold standard.

### **Guideline 4 – Research Contributions**

The fourth guideline states that an effective design science research should provide clear and verifiable contributions in the areas of the design artifact, design foundations, and/or design methodologies. This thesis makes research contributions by designing and developing a system for automating CIRI Human Rights Practice ratings using the GATE platform for natural language processing. It should be noted that this CIRI Automated System is a first of the kind tool developed to automate the CIRI Human Rights Practice ratings process.

### **Guideline 5 – Research Rigor**

The fifth guideline states that a design science research should rely upon the application of rigorous methods in both the construction and evaluation of the design artifact. This thesis utilizes appropriate methods in the construction and evaluation of the automated process. The automated system is built using the GATE platform and subsequently evaluated for F-measure to assess how accurate the automated results is compared to the source data.

**Guideline 6 – Design as a Search Process**

The sixth guideline states that search for an effective artifact requires utilizing available means to reach desired ends while satisfying laws in the problem environment. During the process of designing and developing the automated process various alternatives were considered for application, such as IBM's Watson services as explained in the Chapter 4.

**Guideline 7 – Communication of Research**

The seventh guideline states that a design science research must be presented effectively to both technology-oriented and management-oriented audiences. This thesis follows the communication guideline as the research work is communicated in the form of a written document and an oral presentation as part of thesis proposal and final defense.

## Chapter 4

### GATE TEXT MINING PLATFORM

The General Architecture for Text Engineering or GATE is a 15-year-old open source software project that is used to provide a multitude of solutions for text processing [The University of Sheffield16]. GATE was the software chosen for this research project because of two reasons: 1) the availability of plugins and 2) extensibility via customization.

GATE includes:

- an IDE: GATE Developer
- a web app: GATE Teamware
- a framework: GATE Embedded
- an architecture
- a process for creating services

GATE Developer contains language processing components, an Information Extraction module, and other text processing plugins. GATE Teamware is a collaborative Annotation environment that is used for semantic annotation projects built around a workflow engine and heavily-optimized backend service infrastructure [The University of Sheffield16].

GATE Embedded is used to gain access to the language processing components and plugins by having the GATE resources included within the external application [Cunningham13].

## 4.1 GATE Architecture

Figure 6 is an adapted image of the high-level design of the language processing components of the GATE architecture, taken from GATE's website [The University of Sheffield16]. In the figure, the GATE resources used within this thesis study are highlighted in orange. The GATE process describes the steps needed to create robust and maintainable custom language processing components. The process includes three components: Language Resource, Processing Resource, and Visual Resource. Language Resources (LRs) are entities that hold any type of linguistic data such as report documents, HTML markup, ontologies, etc. Processing Resources (PRs) are entities that process data, for example, extracting entity names and annotating text based on parts of speech. Visual Resource (VRs) are components used for building graphical interfaces such as the panels that display the different types of GATE Annotations after a Process Resource has run [Cunningham13]. Instances of these resources are called CREOLE resources in GATE and contain meta-data that specifies various resource parameters, which are used for initializing a resource instance. These Process Resources can be added to the GATE Controllers, which are used to define GATE applications and control the flow of execution of assigned Process Resources [Cunningham13].

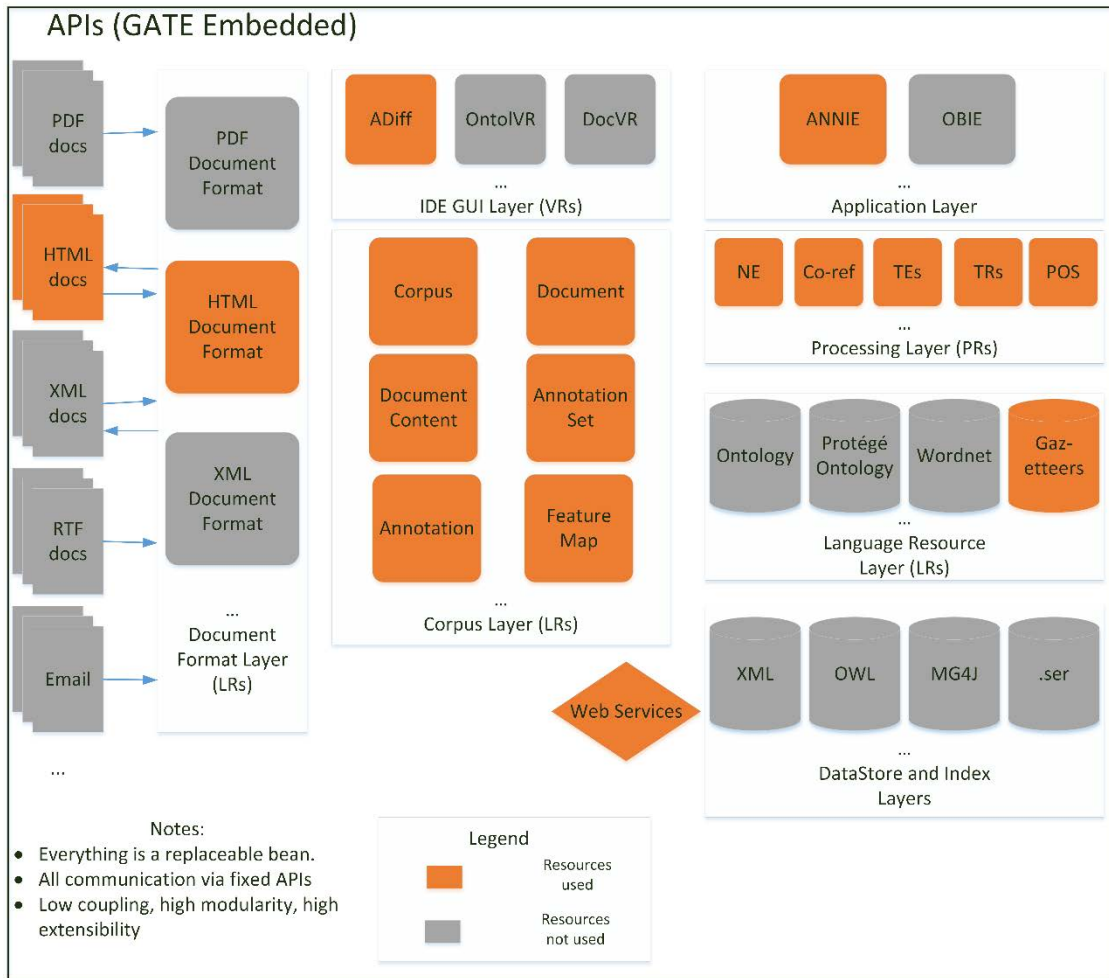


Figure 6. GATE Architecture

#### 4.2 ANNIE: A Nearly-New Information Extraction System

Included in GATE is the Information Extraction system named ANNIE, which is comprised of a tokenizer, a gazetteer, a sentence splitter, a part of speech tagger, semantic tagger, and name tagger as shown in Figure 7 [Cunningham14]. The tokenizer is used for splitting text based on default and custom tokenizer rules. The gazetteer is used to identify entity names in the text based on a set of lists containing data relating to names of cities,

countries, days of the week, units of currency, etc. The sentence splitter is used to separate text into sentences. Sentence splitter uses a gazetteer list to differentiate sentence-marking full stops from other types of symbols. Each sentence and full stop are given the Sentence and Split Annotation, respectively. A part of speech tagger is used to annotate symbols and words. Semantic Tagger, a named entity transducer, is used to produce outputs of annotated entities based on previously assigned annotations similar to the example below. The name tagger (as known as co-reference tagger) module adds identity relations between named entities [Cunningham02].

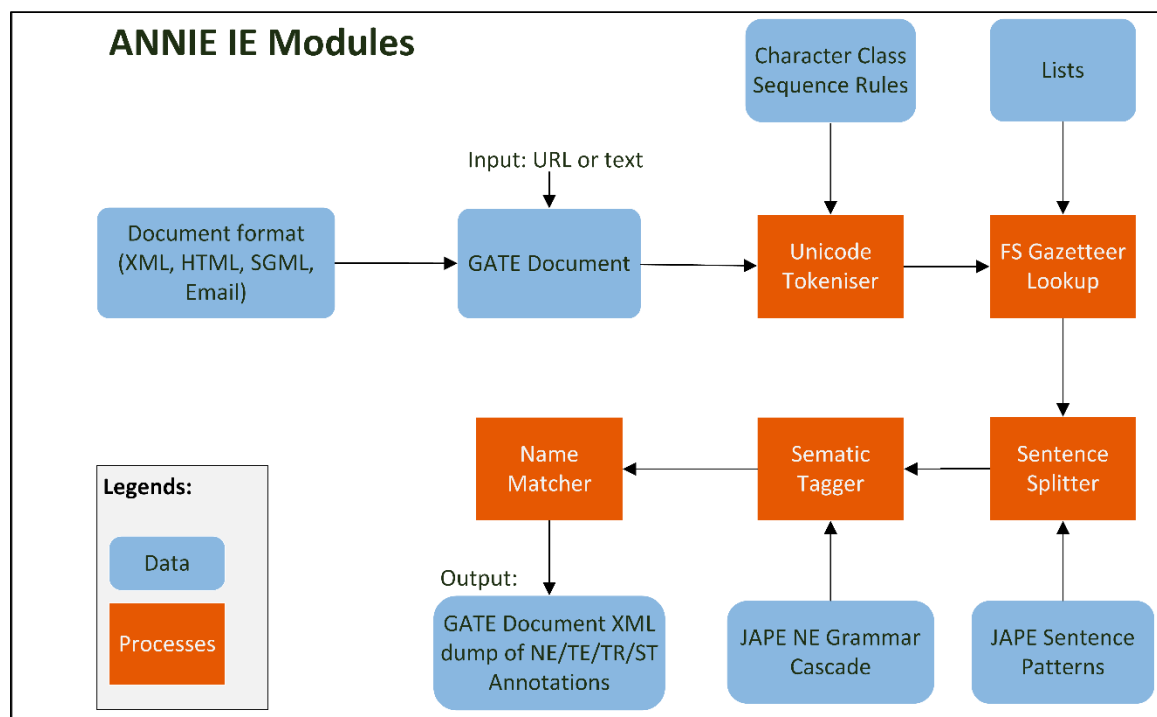


Figure 7. Standard ANNIE Process Flow

The nature of the replaceable modules allows the standard ANNIE process flow to be modified to fit the requirements of this research. In Chapter 5, there is more detail on

which processing resources are used to execute specific functions, but the below figure shows which processing resources were added and which were modified to create a custom ANNIE application. The GATE Morphological analyzer process resource was added to find the root of a word. The Date and Number Tagger process resources were added specific for this research in order to annotate date and number text.

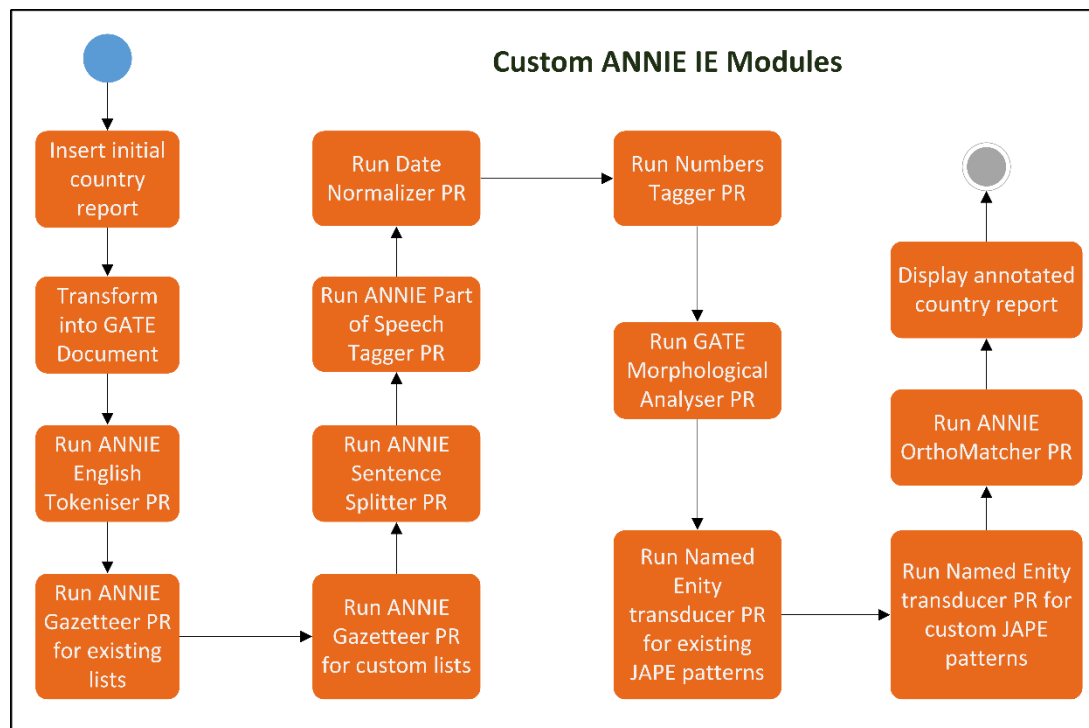


Figure 8. Custom ANNIE Process Flow

ANNIE will be used to extract information and generate Annotations from the United States Department of State Human Rights Country Reports. When a document is imported into GATE as a Language Resource the application converts the displays the document text onto the main resource viewer as plain text. Contained within the main resource viewer is the document editor panel it displays the document text and buttons that help to execute





panel tagged as a certain color. Figure 12 shows the Annotation metadata for each checked Annotation.

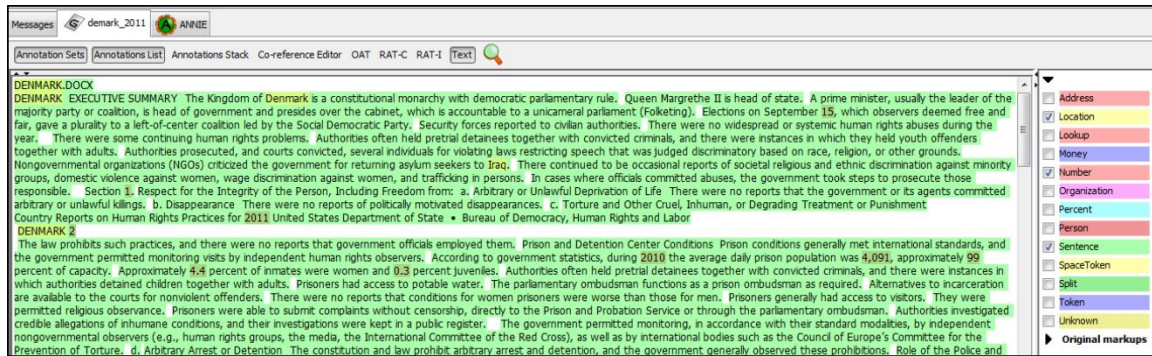


Figure 10. Text tagged with Annotations

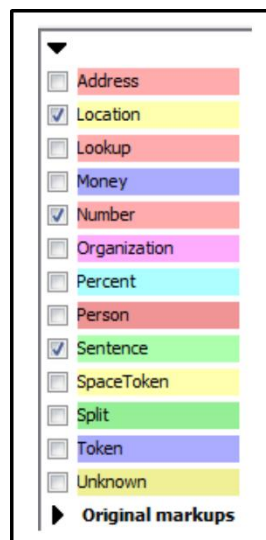


Figure 11. Checked Annotations

Type	Set	Start	End	Id	Features
Location		0	7	33193	{locType=country, matches=[33193, 33194, 33301, 33306, 33321, 33338, 33349, 33361], rule=Location
Sentence		0	12	30395	{}
Location		13	20	33194	{locType=country, matches=[33193, 33194, 33301, 33306, 33321, 33338, 33349, 33361], rule=Location
Sentence		13	128	30396	{}
Location		56	63	33195	{locType=country, matches=[33193, 33194, 33301, 33306, 33321, 33338, 33349, 33361], rule=Location
Sentence		130	166	30397	{}
Sentence		168	354	30398	{}
Sentence		356	503	30399	{}
Number		379	381	33904	{type=numbers, value=15.0}

Figure 12. Annotation List of Annotation metadata

### 4.3 Natural Language Understanding API

Natural Language Understanding is an API developed by IBM that provides text analysis using natural language processing. The API provides functions for analyzing and extracting information from text such as concepts, entities, keywords, sentiment, emotion, and more [IBM17]. Natural Language Understanding is used to more accurately identify the sentiment of selected text by providing a score based on the information extracted from the text, which is further explained in section 5.2. The score has a negative to positive range from -1 to 1. Natural Language Understanding is used to obtain the sentiment score for the various human rights reports sub-sections.

### 4.4 Annotation Diff Tool

The Annotation Diff Tool is a plugin already installed within GATE that enables two Annotation Sets of a document, for example CIRI Annotation and automated Annotation Sets to be compared with each other [Cunningham14]. This tool is used for evaluation

purposes and can compare Annotations of the same type in documents that are either system-generated or hand-annotated. Metrics used by Annotation Diff Tool to assess the extent to which the CIRI Annotation set matches the automated Annotation Sets are precision, recall, and F-measure.

- Precision – measures how many items were correctly identified as a percentage of the number of items retrieved. A high precision rate means that the items being identified are correct.
- Recall – measures how many items were correctly identified as a percentage of the total number of correct items retrieved. A high recall rate means that there is a better chance that items that were correctly identified are not missing.
- F-Measure – is the weighted average of precision and recall.

Precision and recall are calculated by counting occurrences of the following relations:

- Coextensive – if two Annotations are in the same span (start and end offsets are equal) of text within a document.
- Overlaps – if two Annotations share a common span of text.
- Compatible (*Correct*) – if two Annotations are coextensive and have one or more features from the key set included in the Response Set.
- Partially Compatible (*Partial*) – denotes two Annotations are partially correct if they have the same Annotation type and the spans overlap between documents, but one or more features are not identical.

- Missing (*Key Annotations only*) – a Key Annotation is missing if it is not coextensive or overlapping, or if one or more of the features is not included in the response Annotation.
- Spurious (*Response Annotations only*) – a Response Annotation is spurious if it is not coextensive or overlapping, or if one or more of the features is not included in the key Annotation.

Shown below are the standard equations for calculating precision and recall and a visual aid representing *retrieved items*, *relevant items*, *true positives (TP)*, *false positives (FP)*, and *false negatives (FN)*. Figure 13 shows that the total number of *retrieved items* is equal to *true positives* plus *false positives* and *relevant items* is equal to *true positives* plus *false negatives*.

$$Precision = \frac{TP}{TP + FP}$$

Equation 1. Standard Precision Formula

$$Recall = \frac{TP}{TP + FN}$$

Equation 2. Standard Recall Formula

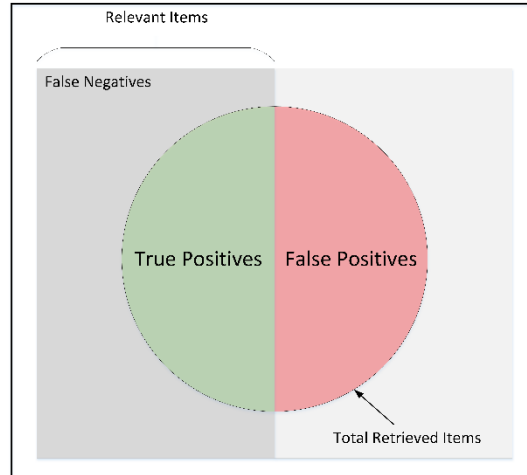


Figure 13. Standard Precision/Recall Diagram

The precision and recall formulas used by the GATE platform are slightly modified from the standard formulas because of the use of Annotations and features. Both precision and recall have the partial items calculated at half weight in the numerator. As mentioned in the Annotation relations above, all of the variables used in the below equations are related to whether the compared Annotations have the same offsets and if the Annotation feature values match. The relation between the standard precision and recall equations and the modified GATE equations is as follows:

- *Correct plus Partial is true positive*
- *Spurious is false positive*
- *Missing is false negative*

With the above relations taken into account that means relevant items and retrieved items are Annotations plus feature values, as shown in Figure 14. Shown below are the modified

equations of precision and recall according to the GATE documentation along with a visual representation of the equations.

$$Precision = \frac{Correct + \frac{1}{2}Partial}{Correct + Partial + Spurious}$$

Equation 3. GATE Precision Formula

$$Recall = \frac{Correct + \frac{1}{2}Partial}{Correct + Partial + Missing}$$

Equation 4. GATE Recall Formula

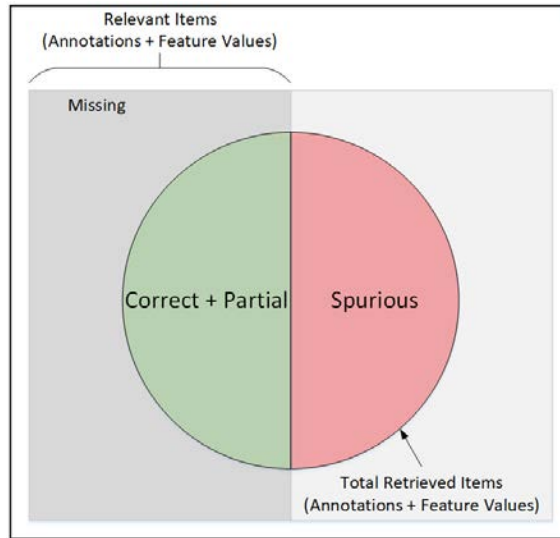


Figure 14. GATE Precision/Recall Diagram

F-measure is the weighted average of precision and recall.  $\beta$  reflects the weight of precision (P) and recall (R). The Annotation Diff Tool allows different weight values to be entered for comparison and a custom weight of 2 will be used in future evaluations.

$$F - measure = \frac{(\beta^2 + 1)P * R}{(\beta^2 P) + R}$$

Equation 5. F-measure

When generated, the values of precision, recall, and F-measure can be also calculated by three criteria: strict, lenient, and average. These criteria are used to deal with partially correct responses in various ways. Cunningham et al., defines these criteria as [Cunningham14]:

- Strict - considers all partially correct responses incorrect. (*spurious*)
- Lenient - considers all partially correct responses correct.
- Average - allocates an average of strict and lenient to weight to partially correct responses.

The Annotation offsets and feature values are used for comparison by comparing the Annotations from the key set (i.e., CIRI dataset) to the Annotations of the Response Set (i.e., automated dataset). First, the Annotation Diff Tool checks the start and end offsets of the Annotations in both documents. If the offsets match, the tool then checks if the features of each Annotation in the Response Set are the same as those in the key set. In the case of this research, the precision and recall scores are the same, meaning the spurious and missing values are the same, due the way the CIRI dataset is structured. The dataset will

always have the same fourteen human rights practices and from years 1999-2011 there is always a rating assigned to a human rights practice. Due to this, the retrieved items, Annotations and feature values will never be empty in either the Key or Response Annotation Sets. Since the retrieved items are not empty, this means the relevant items within the Key and Response Annotation Sets will also never be empty. This essentially means, the Annotation Diff Tool is only comparing whether the feature values within the Annotations of the Key and Response Annotations Set match each other in a one-to-one relationship. Also, because of the relations and metrics defined previously in this section, the only value that changes is the number of correct items based on whether the feature values were missing or spurious. Due to the precision and recall scores being the same, the F-Measure score is used in all evaluation analysis in this thesis. Table 5 shows an example of how the precision and recall are the same based on the above equations.

Year	F-Measure	Key				Response			
		0	1	2	3	0	1	2	3
2000	0.7857	0	1	12	1	1	0	13	0
Precision	Recall	Correct		Partially Correct		Missing		Spurious	
0.7857	0.7857	11		0		3		3	

Table 5. Precision/Recall Example

Denmark, in the table above the precision, recall, and F-Measure metrics have a score of 78.57% for year 2000.



## 4.5 GATE Embedded

The flexibility of the GATE framework comes from being modular. Much of the functionality comes from components or plugins that are loaded manually by the user. The modularity of the GATE framework allows the user to create custom plugins via GATE Embedded that can be used in GATE applications as a Language, Processing, or Visual Resource. The creation of a new resources can be started via the BootStrap Wizard [Cunningham14] in the GATE graphical interface. Using the BootStrap Wizard eliminates the human error by letting the tool create the starting structure. The steps to create a custom process for GATE include:

- write a Java class that implements GATE's beans model;
- compile the class, and any others that it uses, into a Java Archive (JAR) file;
- write some XML configuration data for the new resource; and
- tell GATE the URL of the new JAR and XML files

The GATE Embedded framework is used to create a custom processing resource to produce the automated CIRI ratings. The plugins used are ANNIE, JAPE Plus, Gazetteer, Date Tagger, Number Tagger, and Natural Language Understanding, all of which can be used to create a custom Process Resource. New Gazetteer lists (see Appendix A) and JAPE Grammar rules (see Appendix B) are created for identifying various patterns of dates, currencies, numbers, country report sections as well as keywords and phrases of interest within the document. Customized JAPE Grammar rules are processed along with ANNIE in order to identify patterns that apply to CIRI Identifier.

## Chapter 5

### IMPLEMENTATION

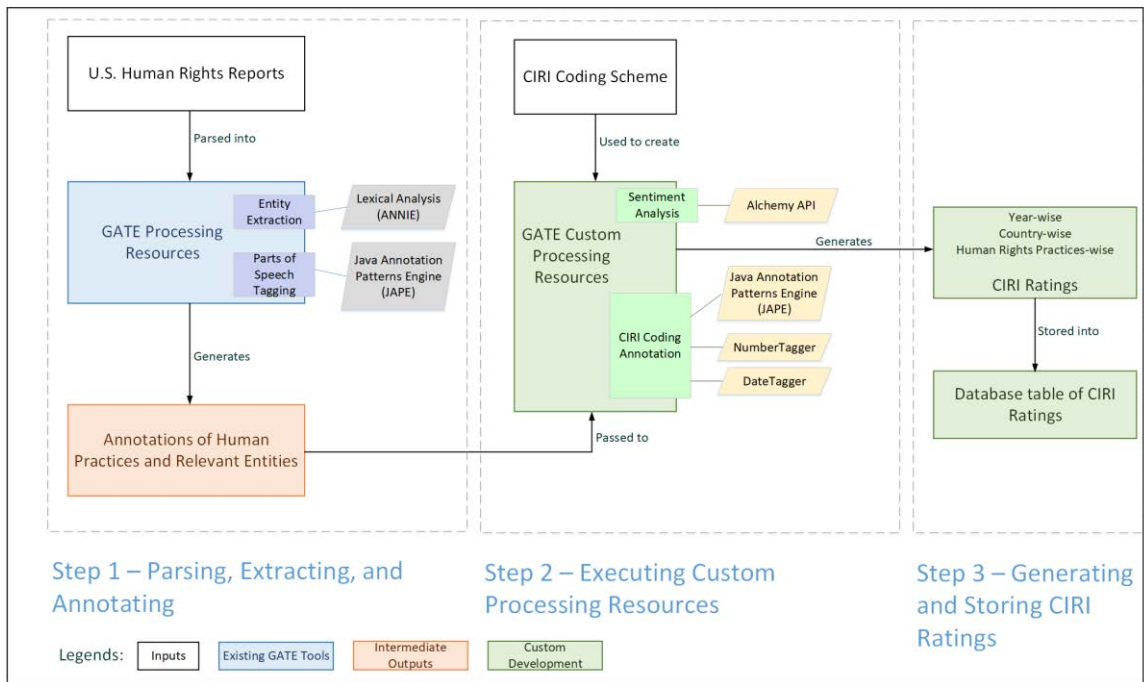


Figure 15. Automated Development Process

GATE is an Annotation based tool and relies on the correct document structure, patterns, rules, and relations to give accurate data. To get started, a GATE application is created that consists of Process Resources, Language Resources, and Visual Resources. The Processing Resources include JAPE Plus, Date Tagger, Number Tagger, ANNIE, and Natural Language Understanding. The objective is to extract information from the subsections of the United States Human Rights Reports and return a rating value for each subsection. The result of each subsection value should correspond to the values listed for

each Human Practice Code in the CIRI dataset. As shown in Figure 15, there are three main steps within the application:

1. Parsing the text and annotating the relevant entities in the report document with ANNIE.
2. Using the generated Annotations from ANNIE with the CIRI coding schemes to create the custom processing resource.
3. Generating the automated CIRI ratings and storing the values into a MySQL database table for each country and year. The stored data will then be retrieved to evaluate the accuracy of the automatically generated values with those listed in the CIRI dataset.

The structure of the GATE application can be seen below in Figure 16. This application is saved and then loaded into the custom processing resource so that these existing processing resources can be called every time a country report needs annotating during Step Two.

Annotations generated in Step Two is stored in a database table during Step Three. Details of above three steps are explained in the following sub-sections.











Selected Processing resources		
!	Name	Type
	Document Reset PR	Document Reset PR
	ANNIE English Tokeniser	ANNIE English Tokeniser
	ANNIE Gazetteer	ANNIE Gazetteer
	ANNIE Sentence Splitter	ANNIE Sentence Splitter
	ANNIE POS Tagger	ANNIE POS Tagger
	Date Normalizer 00015	Date Normalizer
	Numbers Tagger 00013	Numbers Tagger
	GATE Morphological analyser 00016	GATE Morphological analyser
	ANNIE NE Transducer	ANNIE NE Transducer
	ANNIE OrthoMatcher	ANNIE OrthoMatcher

Figure 16. GATE Application PR Structure

## 5.1 Step One – Parsing, Extracting, and Annotating

The setup process for Step One includes crawling the United States Department website for all the available Country Reports and identifying the required GATE plugins to run processes over the reports. The Open Source web crawler called Jsoup [Hedley17] is used to retrieve the Country Reports from 1999-2011 from the US Department website. Jsoup is used to extract the relevant data and exclude any extraneous HTML markup. The reports are saved as HTML files and then placed into the pipeline of the custom processing resource. These country report documents can be processed to generate Annotations that only cover the country report sections defined in custom JAPE grammars, which is explained in the next section.

## 5.2 Step Two – Executing Custom Processing Resources

Step Two uses JAPE Plus patterns to create annotation based on document structures observed within the Human Rights country reports. The annotations are created based on the section headings and numerical values as delimiters. The annotations for the report sections are created by listing each section within the ANNIE Gazetteer, which when run generates the Annotation Set consisting of the country report section annotations. Lists of items, such as section headers or keywords, are created by modifying the Gazetteer text files because it provides more flexibility with customizing the Gazetteer to include the CIRI coding schemes. These text files allow the country report section annotations to be mapped to their respective features [Cunningham14].

Annotations based on numerical values found within the text are used to find the number of occurrences of Human Practice violations. The CIRI Coding Manual states which country report sections use the numerical frequency coding scheme shown in Table 2 [Cingranelli14B]. A CIRI rating of 0 is high, 1 is moderate, and 2 is low. High violations are occurrences of 50 or more, moderate violations are those between 1 and 49, and low violations are 0 or no occurrences reported [Cingranelli14B]. The Number Tagger is used to create Annotations of all the numbers in the document. The value of the entities found within the Number Annotations are then used to place a CIRI rating on the particular section. An issue did arise involving some instances of date or currency being picked up as Numbers, so custom JAPE grammars were created that excluded those types of numbers from being included in the generation of the custom Number Annotation.

The custom Process Resource takes the output Annotation Set and searches for the Lookup, Ciri\_SubLevel1, Ciri\_SubLevel2, and Ciri\_SubLevel3 Annotations to extract the delimiters and then take the content between starting and ending delimiters to create four new Annotations based on the country report subsections, SectionLevel0 Content, SectionLevel1 Content, SectionLevel2 Content, and SectionLevel3 Content. With the new Annotations created, the text of the Annotations is retrieved and checked to determine if the report section headers are contained within the text. If the check succeeds, the GATE document, the Annotation, and the Output Annotation Set are passed to three types of methods for further processing.

These three main method types are used for calculating the rating value for each section: 1) quantitative, 2) qualitative, and 3) women's rights. Each of these types were taken from the CIRI manual where it states the type of scores given to each Human Rights Practice. The practices related to killings, disappearances, political prisoners, and torture were given scores from 0 to 2 that were based on the occurrence of human practice violations. All other practices except for Women's Rights, which is on a 0 to 3 scale, were also score on a 0 to 2 scale, but relied more on keywords such as "not restricted", "restricted", or "severely restricted". Keywords and key phrases, based on CIRI coding schemes, help to create an accurate CIRI rating, but because some sections within the country reports have ambiguous wording or inconsistent scoring from the source CIRI dataset context of the text is relied on for additional accuracy checks. Including Natural Language Understanding into the

custom process helps analyze the context of the text better by using the Sentiment Analysis service to generate a sentiment score for portions of text.

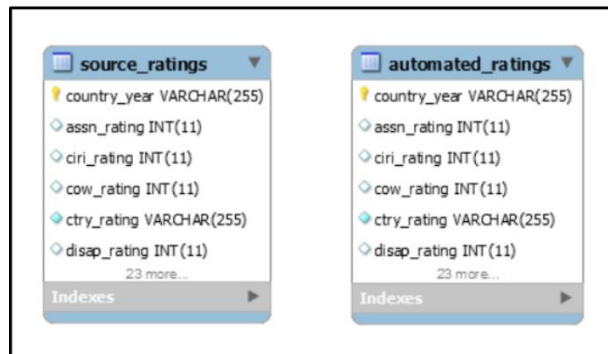
In the quantitative method, the CIRI Annotations are searched for each Annotation instance, the country report section, and the number of Sentence Annotations contained within the Annotation instance. Based on the patterns of the Human Rights Country Reports, sections with only one sentence have no human rights practice violations and therefore a rating value of 2 is given. Keywords such as “no reports” gets a rating value of 2 unless the sentence contains keywords such as “occasion”, “credible reports”, or “numerous” that changes the value to 1 [Cingranelli14B]. The keyword “were reports” gets a rating value of 0 unless combined with some of the other keywords that will make the rating value equal 1, such as words that state whether the actions were carried out by government or government related entities. The numeric values found, such as number of killings or disappearances, when searching the country report sections are checked only if the sentence contains text that mentions there were at least some reports of violations. The number of occurrences found within the text can change the rating value from 0 to 1 or vice versa. If no rating value was assigned during this process then the sentence strings are passed to the Natural Language Understanding API to analyze the sentiment of the text and return a sentiment score. The sentiment score ranges from -1(negative) to 1(positive). If the sentiment score is positive (0.00 and above) then the rating value is set to 0. Negative sentiment scores are not used for assigning rating values because of a high variance due to the frequency of words that the sentiment tool deems negative.

The qualitative method type is much the same as the quantitative type except that it doesn't have a numeric value to check for occurrences of human practice violations. The keywords in this type of method includes words such as, "government restrictions", "law prohibits", "limit free speech", "did not respect", or "respected".

As with the quantitative section, if no coding scheme patterns are found within the country report sections the text is passed to a method that calls the Sentiment Analysis service, in the Natural Language Understanding API, and returns a sentiment score between -1(negative) and 1(positive) with two significant digits. The IBM documentation calls this type of service Document Sentiment where the attributes are sentiment and score; sentiment is labeled positive or negative and score is the numerical value [IBM17]. When the sentiment score is 0.0 then the rating value is set to 2 because the text doesn't contain enough information to generate any other rating value or 3 when concerning sections related to human practice ratings for the Women. If the sentiment score is positive then the rating value is also set to 2, or 3 if the text processed is for any of the Women's human rights sections. Negative sentiment scores are ignored due the amount of variance between a country report section that has a rating of 0 or 1 and some false negatives when the text processed should be rated as a 2.



### 5.3 Step Three – Generating and Storing CIRI Ratings



The image shows two side-by-side screenshots of a MySQL database interface. The left screenshot displays the structure of the 'source\_ratings' table, and the right screenshot displays the structure of the 'automated\_ratings' table. Both tables have identical column definitions: 'country\_year' (VARCHAR(255)), 'assn\_rating' (INT(11)), 'ciri\_rating' (INT(11)), 'cow\_rating' (INT(11)), 'ctry\_rating' (VARCHAR(255)), and 'disap\_rating' (INT(11)). Each table also has an 'Indexes' section at the bottom, which is currently collapsed, showing '23 more...'. The tables are presented in a light blue-themed window.

Table Name	Column Name	Column Type
source_ratings	country_year	VARCHAR(255)
	assn_rating	INT(11)
	ciri_rating	INT(11)
	cow_rating	INT(11)
	ctry_rating	VARCHAR(255)
	disap_rating	INT(11)
automated_ratings	country_year	VARCHAR(255)
	assn_rating	INT(11)
	ciri_rating	INT(11)
	cow_rating	INT(11)
	ctry_rating	VARCHAR(255)
	disap_rating	INT(11)

Figure 17. CIRI Database Table Structure

Once Step Two of the Automated Development Process is complete, the last step of the process is to generate the human rights practice ratings and store the generated rating values in the `automated_ratings` database table. As seen in Figure 17, the MySQL database contains two similarly structured tables: `source_ratings` and `automated_ratings`. The purpose of the `source_ratings` table will be described in Chapter 6, Evaluation. The generated rating values are put into a key-value map where key is the human right practice indicator and value is the practice rating. Once all of the ratings are generated the map is passed to the database access class where the map is iterated over and the values are stored into the `automated_ratings` table. A `country_year` (e.g. `Canada_2011`) is used as the primary key in the table and each indicator of the CIRI dataset has its own column denoted as `practice_rating` (e.g. `disap_rating`).

The entire process for steps two and three took an average of about one hour and 30 minutes to annotate and store 20 country reports. The process of reading the text and

storing human practices and ratings were fast, but the bottleneck of this process was the ANNIE component due to the amount of annotations occurring in the text.

## Chapter 6

### EVALUATION

The objective of this evaluation is to measure accuracy of the results produced by the automated process in comparison with the source CIRI Human Rights Practice ratings. Accuracy is determined by the results produced by the precision, recall, and F-measure metrics. These results will be calculated using GATE's Annotation Diff Tool referenced in section 4.4, but as mentioned in that section the F-Measure score will be the evaluation measure used for analysis. Measuring precision, recall, and F-Measure for the Human Rights Country reports is important in this research because the main text mining technique used in the automated process is text classification.

Our choice of assessment metrics are similar to other studies that evaluated text classification techniques. In the paper, "Multilabel Text Classification for Automated Tag Suggestion", the authors use precision, recall, and F-measure as an appropriate evaluation method for tuning parameters and evaluating frameworks in their automated recommender system [Katakis08]. They used three training sets consisting of tags between 50 to 300 and words between 100 and 3000 to test predictions and achieved results from the F-measure above 70% for all three sets [Katakis08]. In the survey titled, "A Brief Survey of Text Mining", the authors use F-measure to calculate the accurateness of an automated text classification system [Hotho05]. In the text classification application conducted, they found that a corpus with 2300 categories achieved an F-measure score of 39%, while a

corpus of 800 categories achieved a score of 79% with partial automation [Hotho05]. The latter case uses human categorizers in conjunction with the automated system because the text classification provides more consistency and faster annotation [Hotho05].

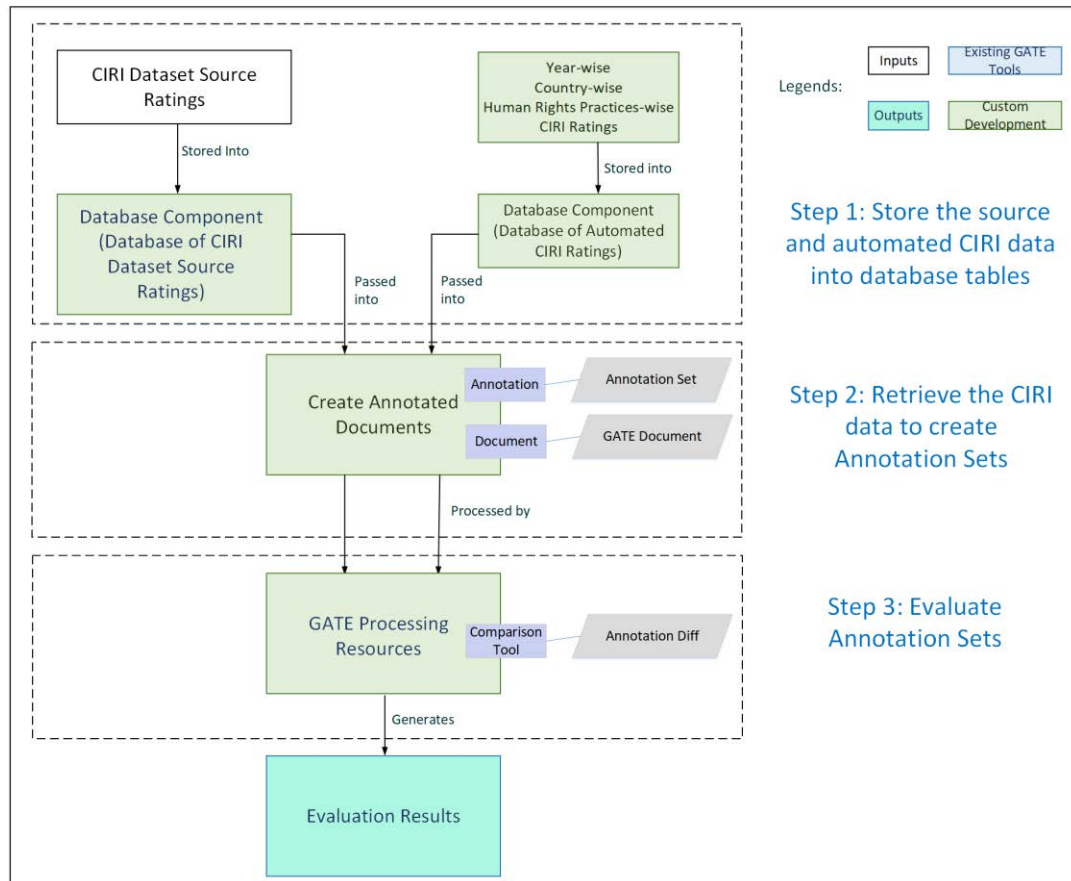


Figure 18. Evaluation Plan

Figure 18, above, consists of three steps: 1) store the source and automatically annotated CIRI Human Rights Practices and ratings into their respective source\_ratings and automated\_ratings table, 2) retrieve CIRI data to create Annotation Sets, and 3) evaluate the Annotation Sets using the Annotation Diff interface from the Gate Embedded framework.

## 6.1 Step One – Store the Source and Automated CIRI Data into Tables

The first step consists of reading in data from the CIRI dataset to get the human rights practice ratings and putting these values into a key-value map where key is the human rights practice indicator and value is the practice rating. This map is passed to the database access class within the Database Component where it is iterated over and stored into the `source_ratings` table (See Figure 17 for the table structure).

## 6.2 Step Two – Retrieve CIRI Data to Create Annotation Sets

A new Processing Resource is created for Step Two to retrieve the human rights practices and its rating values from both the source and automated CIRI tables. The data from these tables are retrieved from the database access class that exists within the Database Component. The human rights practices and rating values extracted from the source table is used as the Key Set and the data from the automated table is used as the Response Set for evaluation.

A GATE Document is created for each country-year from the list of maps returned from the Database Component. These GATE documents are used to create and populate five types of Annotation Sets. One Annotation Set contains the evaluation data for every human practice right, another gathers data specific to the Women's human practice rights, two Annotations Sets are for data based on the Physical Integrity Index and Empowerment Index groups, and the last is for Independent Judiciary. These Annotation Sets contain

annotation instances which denote the human practice rights for a given country and year by adding the human practice ratings names and their rating values as features (key-value) to the instance.

### 6.3 Step Three – Evaluate the Annotation Sets

There are two ways to use the Annotation Diff Tool, either manually via the GATE GUI as shown below in Figure 19, or by using the GATE Embedded framework for customized development. The Annotation Diff Tool compares a Key Set and Response Set for Annotation Sets with an option to include all, some, or none of the Annotation features. The GATE GUI is used by selecting the desired Key and Response sets and target Annotation Type, denoted as country\_year, from the Key and Response documents. Once these respective dropdowns have the desired items selected, the Compare button is pressed to begin the evaluation. This research instead uses the GATE Embedded framework to create custom code that automates the GATE GUI process by evaluating all the Annotation Sets at once. The evaluation results are stored as HTML files.

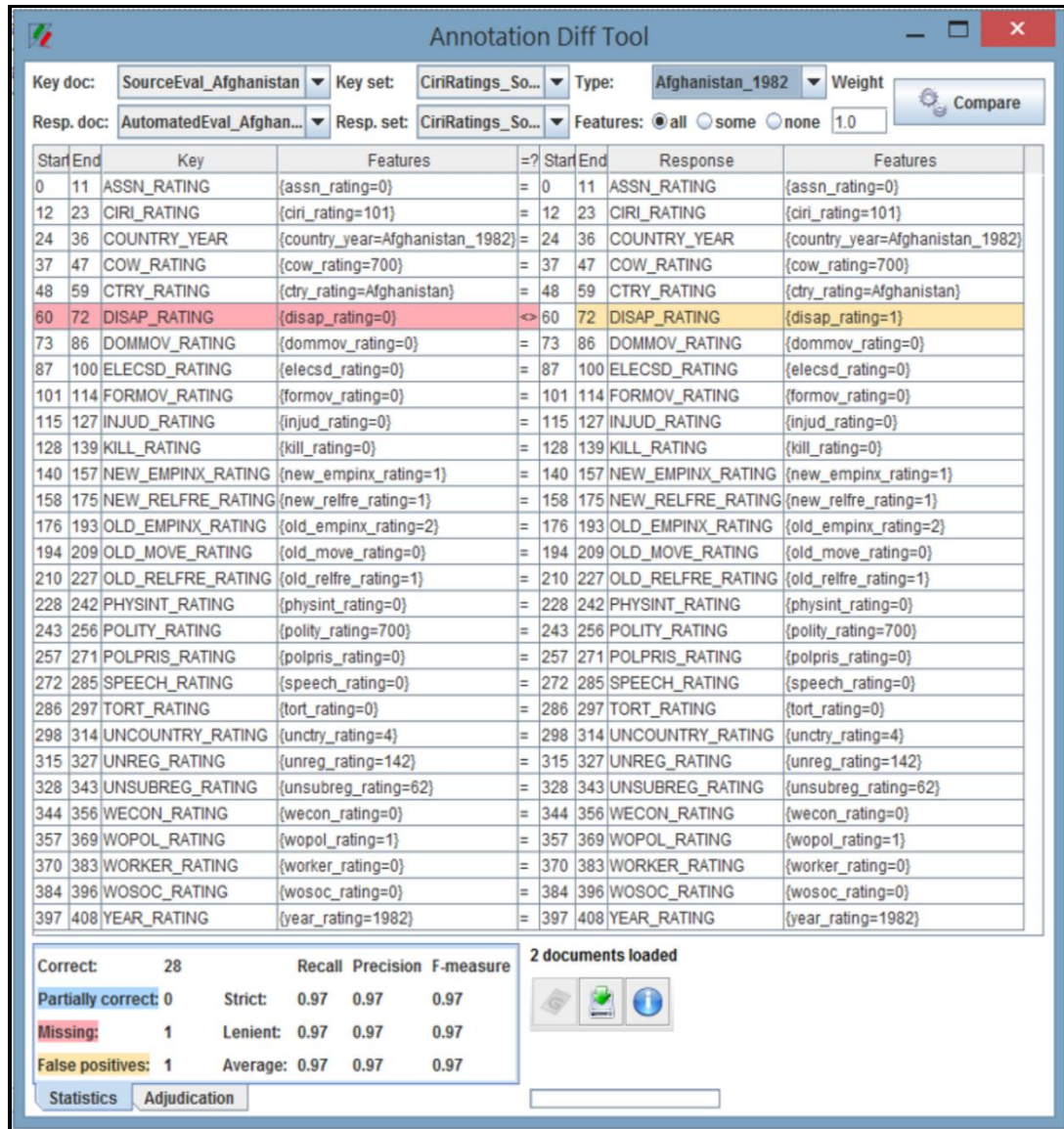


Figure 19. CIRI Document Comparison

## 6.4 Data Collection

The evaluation results stored in the HTML files contain two types of data: evaluation measures and CIRI ratings. Each type of data is extracted from the HTML file and stored into its own “measures” and “ratings” CSV files. These CSV files contain data sorted by

earliest year first for easier storage, readability, and data manipulation. The “measures” file contain GATE measures (such as F-Measure, Correct, Partially Correct, Missing, and False Positive) and number of occurrences for each human practice rating value (such as 0, 1, 2, or 3) for each country per year in the Key and Response sets. The “ratings” files include the evaluation results for the individual human practice ratings for each country per year pertaining to whether the human practice ratings for the source Annotation feature matched the value for the automated Annotation feature.

From Table 1, it can be noted that there are fourteen human rights practices, which are grouped into four categories: Independent Judiciary, Physical Integrity, Empowerment Rights, and Women's Rights. Below, we provide trends and analysis for these four categories.

## 6.5 Trends and Analysis

In regards to preparing the dataset for further analysis, we identified and removed outliers. Outliers in this context would be periods of interruption or interregnum that occur in certain countries. For example, during 2003 and 2004 Afghanistan was in a state of war (periods of interruption), CIRI assigns a rating value of -66 for all human practice ratings even though the United States published a country report for Afghanistan during these years. Similarly, CIRI assigns a rating value of -77 when a country has no central governmental authority for a given year. In our CIRI Automated System, we followed regular coding scheme to produce human right scores for those years. As a result, F-Measure score for



those years would be 0 or close to 0, which are considered as outliers and are removed from the dataset for trends and analysis.

CIRI rated human rights practices for 204 countries for the years 1981 to 2011. CIRI Automated System rated human rights practices for 204 countries for the years 1999 to 2011, as explained in Chapter 1. F-measure scores were generated from the GATE annotation diff tool for the years 1999 to 2011. The entire evaluation results for all countries can be viewed separately by downloading the CSVs at the following website: <https://cirithesissite.herokuapp.com>. Given limited space in this thesis and for reasons of providing meaningful presentation, we will demonstrate trends and analysis for four human rights practice categories for a few selected countries. We vary the selection of countries used for demonstration across four categories to provide appropriate representation of 204 countries. Table 6 provides listing of twelve sample countries used for demonstrating trends and analysis. The United States Human Rights Reports have countries split into six regional groups: Africa, East Asia and the Pacific, Europe and Eurasia, Near East and North Africa, South and Central Asia, and Western Hemisphere. The names of these groups may vary depending on the year of the reports, but the countries are the same unless they did not exist during the time period. Two countries were chosen from each group to make up the twelve sample countries used. Summarized evaluation results for these twelve countries can be found in Appendix C.

<b>Region Group</b>	<b>Country</b>
Africa	Angola, Sudan
East Asia and the Pacific	Australia, Cambodia
Europe and Eurasia	Austria, Denmark
Near East and North Africa	Bahrain, Egypt
South and Central Asia	Bhutan, Pakistan
Western Hemisphere	Canada, Colombia

Table 6. Evaluation Sample Countries

#### 6.5.1 Four Human Rights Practice Groups

Of the data gathered from 204 countries, the overall trend observed was that the CIRI Automated System did better at rating human rights practices when the patterns were based on less ambiguous coding schemes. The F-Measure for the Physical Integrity group was the most accurate out of all of the human rights practices groups (Independent Judiciary, Physical Integrity, Empowerment Rights, and Women's Rights). The defining difference is that the coding scheme for Physical Integrity group was based on numerical amounts instead of keywords or phrases. See Tables 7 to 11 for Denmark's evaluation results.

Each Human Rights Practice Group affects the overall F-Measure score during evaluation calculations. The precision and recall scores, as defined in Chapter 5, are affected by the number of human rights practice indicators within the Key and Response Sets during evaluation. The Independent Judiciary group has one of the fourteen human rights practices, the Physical Integrity group contains four of the fourteen human rights practices, the Empowerment Rights group has the largest with seven human rights practices, and the Women's rights group, which contains two of the fourteen human rights practices. The

number each individual group affects the precision and recall scores, which in turn affects the F-Measure score. In essence, the more human rights practices a group has, then it will indirectly, through the precision and recall scores, have a larger effect on the overall F-Measure score. Moreover, human rights practice groups with a higher number of qualitative based criteria will negatively affect the F-Measure score when compared to groups with quantitative based criteria.

The Independent Judiciary, Empowerment Rights, and Women's Rights practices are qualitative, which means the CIRI Automated System may not detect context of keywords or phrases. In cases when CIRI has a highly ambiguous coding scheme, the CIRI Automated System could not accurately detect text patterns associated with those human rights practices, thus, resulting in discrepancies between key and response sets, which lead to lower F-Measure scores.

Tables 7 through 10 show the F-Measure and Key and Response ratings between years 1999 and 2011 for the four groups (Independent Judiciary, Physical Integrity, Empowerment Rights, and Women's Rights, respectively) for Denmark. These tables display the number of times the ratings of 0 – 3 occurred in the Key and Response. Table 11 shows the cumulative F-Measure score of all of the human rights practices from the above mentioned groups. In Table 10, the F-Measure for Women's Rights are scored the worst for Denmark with the lowest F-Measure score is 0% and the highest is 50% compared to the other groups. Table 11 shows that between years 2007 and 2010 the

F-Measure score is between 50% - 57.14%. For these particular years, the F-Measure scores are lower in comparison to other years. Closer observation reveals that there are more occurrences of 1 or 3 ratings for those years, which indicates that the CIRI automated tool did not recognize the required patterns for rating those sections in the country reports. Tables 9 and 10 for Empowerment Rights and Women's Rights, respectively, show a high correlation with the overall results in Table 11 for years 2007 to 2010 where the F-Measure score is low.

Year	F-Measure	Key			Response		
		0	1	2	0	1	2
1999	1	0	0	1	0	0	1
2000	1	0	0	1	0	0	1
2001	1	0	0	1	0	0	1
2002	1	0	0	1	0	0	1
2003	1	0	0	1	0	0	1
2004	1	0	0	1	0	0	1
2005	1	0	0	1	0	0	1
2006	1	0	0	1	0	0	1
2007	1	0	0	1	0	0	1
2008	1	0	0	1	0	0	1
2009	1	0	0	1	0	0	1
2010	1	0	0	1	0	0	1
2011	1	0	0	1	0	0	1

Table 7. Denmark Independent Judiciary Results

Year	F-Measure	Key			Response		
		0	1	2	0	1	2
1999	1	0	0	4	0	0	4
2000	0.75	0	1	3	0	0	4
2001	1	0	0	4	0	0	4
2002	1	0	0	4	0	0	4
2003	0.75	0	0	4	0	1	3
2004	1	0	0	4	0	0	4
2005	1	0	0	4	0	0	4
2006	1	0	0	4	0	0	4
2007	0.75	1	0	3	0	0	4
2008	0.75	0	0	4	0	1	3
2009	0.75	0	1	3	0	0	4
2010	0.75	0	0	4	0	1	3
2011	1	0	0	4	0	0	4

Table 8. Denmark Physical Integrity Results

Year	F-Measure	Key			Response		
		0	1	2	0	1	2
1999	0.8571	0	0	7	1	0	6
2000	0.8571	0	0	7	1	0	6
2001	0.5714	0	0	7	3	0	4
2002	0.8571	0	0	7	1	0	6
2003	0.8571	0	0	7	1	0	6
2004	0.7143	0	0	7	0	2	5
2005	0.5714	0	1	6	0	4	3
2006	0.4286	0	0	7	1	3	3
2007	0.4286	0	2	5	1	4	2
2008	0.5714	0	2	5	0	5	2
2009	0.5714	0	2	5	0	5	2
2010	0.5714	0	1	6	0	4	3
2011	0.8571	0	2	5	1	1	5

Table 9. Denmark Empowerment Rights Results

Year	F-Measure	Key				Response			
		0	1	2	3	0	1	2	3
1999	0.5	0	0	1	1	0	0	2	0
2000	0.5	0	0	1	1	0	0	2	0
2001	0.5	0	0	1	1	0	0	2	0
2002	0.5	0	0	1	1	0	0	2	0
2003	0.5	0	0	1	1	0	0	2	0
2004	0	0	0	0	2	0	2	0	0
2005	0	0	0	1	1	0	2	0	0
2006	0.5	0	1	0	1	0	2	0	0
2007	0	0	0	1	1	0	2	0	0
2008	0	0	0	0	2	0	2	0	0
2009	0	0	0	0	2	0	0	0	0
2010	0	0	0	0	2	2	0	0	0
2011	0	0	0	0	2	0	2	0	0

Table 10. Denmark Women's Rights Results

Year	F-Measure	Key				Response			
		0	1	2	3	0	1	2	3
1999	0.8571	0	0	13	1	1	0	13	0
2000	0.7857	0	1	12	1	1	0	13	0
2001	0.7143	0	0	13	1	3	0	11	0
2002	0.8571	0	0	13	1	1	0	13	0
2003	0.7857	0	0	13	1	1	1	12	0
2004	0.7143	0	0	12	2	0	4	10	0
2005	0.6429	0	1	12	1	0	6	8	0
2006	0.6429	0	1	12	1	1	5	8	0
2007	0.5	1	2	10	1	1	6	7	0
2008	0.5714	0	2	10	2	0	8	6	0
2009	0.5714	0	3	9	2	0	7	7	0
2010	0.5714	0	1	11	2	2	5	7	0
2011	0.7857	0	2	10	2	1	3	10	0

Table 11. Denmark Overall Evaluation Results

### 6.5.2 Independent Judiciary

Tables 12 to 14 below show the human rights practice rating value for the Key and Response Set for years 1999-2011 for Canada, Cambodia, and Bahrain. From Tables 12 and 13, it can be noted that for all country reports analyzed for Canada and Cambodia, 100% F-Measure score was achieved. However, for Bahrain (see Table 14), the country reports analyzed received only 7.7% F-Measure score as only year 2007 key and response sets were matched.

Year	Independence of the Judiciary	
	Key	Response
1999	2	2
2000	2	2
2001	2	2
2002	2	2
2003	2	2
2004	2	2
2005	2	2
2006	2	2
2007	2	2
2008	2	2
2009	2	2
2010	2	2
2011	2	2

Table 12. Canada Independent Judiciary Key and Response Data

Year	Independence of the Judiciary	
	Key	Response
1999	0	0
2000	0	0
2001	0	0
2002	0	0
2003	0	0
2004	0	0
2005	0	0
2006	0	0
2007	0	0
2008	0	0
2009	0	0
2010	0	0
2011	0	0

Table 13. Cambodia Independent Judiciary Key and Response Data

Year	Independence of the Judiciary	
	Key	Response
1999	1	0
2000	1	0
2001	0	2
2002	0	2
2003	0	2
2004	0	2
2005	0	2
2006	0	2
2007	0	0
2008	1	2
2009	1	2
2010	1	2
2011	1	2

Table 14. Bahrain Independent Judiciary Key and Response Data



### 6.5.3 Physical Integrity

Tables 15 to 20 below show information regarding Angola, Sudan, and Colombia, respectively, across thirteen years for the Physical Integrity group. Tables 15, 17, and 19 show a more detailed view of the Key and Response Set for each human rights practice rating value. Tables 16, 18, and 20 show the associated F-Measure scores for the Physical Integrity group. Angola has an F-Measure score at or above 75% for four out of thirteen years, Sudan has twelve years with an F-Measure score above 75%, and Colombia has an F-Measure score at or above 75% for all years. Of these three countries, Colombia is the most accurate in this group with seven 100% F-Measure scores while Sudan has six 100% F-Measure scores and Angola has one 100% F-Measure score. The average Physical Integrity F-Measure score for Sudan is 78.57%, Angola has an average F-Measure of 50%, and Colombia's F-Measure score is 88.46%.

The above data show that the F-Measure score was generally lower when occurrences with ratings of 1 appeared more frequently in the CIRI dataset. The least accurate human rights practice rating was Political Prisoners and the most accurate was Killings. These sample countries have an F-Measure score at 75% or above for most years. Even though the Physical Integrity group provides the most accurate scores for the CIRI Automated System it has the second highest effect on the overall evaluation scores because of the number of human rights practices within the group.

Year	Disappearances		Political Prisoners		Killings		Torture	
	Key	Response	Key	Response	Key	Response	Key	Response
1999	0	0	2	0	0	0	0	0
2000	0	1	1	0	0	0	0	0
2001	0	0	1	0	0	0	0	0
2002	0	0	2	0	0	0	0	0
2003	1	2	2	0	0	0	0	0
2004	1	2	2	0	0	0	0	0
2005	1	2	1	0	1	0	0	0
2006	2	2	1	1	1	1	0	0
2007	1	2	1	0	0	0	0	0
2008	1	2	1	0	1	0	0	0
2009	1	2	0	0	1	0	0	0
2010	2	2	1	0	1	0	0	1
2011	1	2	1	0	1	0	1	0

Table 15. Angola Physical Integrity Key and Response Data

Year	F-Measure	Correct	Missing
1999	0.75	3	1
2000	0.50	2	2
2001	0.75	3	1
2002	0.75	3	1
2003	0.50	2	2
2004	0.50	2	2
2005	0.25	1	3
2006	1	4	0
2007	0.50	2	2
2008	0.25	1	3
2009	0.50	2	2
2010	0.25	1	3
2011	0	0	4

Table 16. Angola Physical Integrity F-Measure Scores

Year	Disappearances		Political Prisoners		Killings		Torture	
	Key	Response	Key	Response	Key	Response	Key	Response
1999	0	0	0	0	0	0	0	0
2000	0	0	0	0	0	0	0	0
2001	0	0	0	0	0	0	0	0
2002	0	0	0	0	0	0	0	0
2003	0	0	0	0	0	0	0	0
2004	2	0	0	0	0	0	0	0
2005	1	0	0	0	0	0	0	0
2006	1	0	0	0	0	0	0	0
2007	1	0	0	0	0	0	0	0
2008	0	0	0	0	0	0	0	0
2009	1	0	0	0	0	1	0	0
2010	1	0	0	0	0	0	0	0
2011	1	0	0	0	0	0	0	0

Table 17. Sudan Physical Integrity Key and Response Data

Year	F-Measure	Correct	Missing
1999	1	4	0
2000	1	4	0
2001	1	4	0
2002	1	4	0
2003	1	4	0
2004	0.75	3	1
2005	0.75	3	1
2006	0.75	3	1
2007	0.75	3	1
2008	1	4	0
2009	0.50	2	2
2010	0.75	3	1
2011	0.75	3	1

Table 18. Sudan Physical Integrity F-Measure Scores

Year	Disappearances		Political Prisoners		Killings		Torture	
	Key	Response	Key	Response	Key	Response	Key	Response
1999	0	0	0	0	1	0	0	0
2000	0	0	0	0	0	0	0	0
2001	0	0	1	0	0	0	0	0
2002	0	0	2	0	0	0	0	0
2003	0	0	0	0	0	0	0	0
2004	0	0	0	0	0	0	0	0
2005	0	0	0	0	0	0	0	0
2006	0	0	0	0	0	0	0	0
2007	0	0	0	0	0	0	0	0
2008	0	0	0	0	0	0	0	0
2009	0	0	0	0	0	0	1	0
2010	0	0	0	0	0	1	1	0
2011	0	0	0	0	0	0	1	0

Table 19. Colombia Physical Integrity Key and Response Data

Year	F-Measure	Correct	Missing
1999	0.75	3	1
2000	1	4	0
2001	0.75	3	1
2002	0.75	3	1
2003	1	4	0
2004	1	4	0
2005	1	4	0
2006	1	4	0
2007	1	4	0
2008	1	4	0
2009	0.75	3	1
2010	0.75	3	1
2011	0.75	3	1

Table 20. Colombia Physical Integrity F-Measure Scores

#### 6.5.4 Empowerment Rights

The Empowerment Rights group is different from other groups in that it has the most human rights practices that does not have numerical coding schemes, and has a rating range from 0 to 2. The trend observed among seven human rights practice indicators within the Empowerment Rights group shows that countries that received higher number of 2 ratings are more accurate than those countries with higher number of 1 ratings. The overall F-Measure score (including all four groups) mostly depends on this group the most because of the qualitative nature of the coding schemes for these particular human rights practices and the amount of human practices contained within the group (seven).

The F-Measure for this group for Austria has ten years with F-Measure scores below 50% and three years with score at 57% and an F-Measure score average of 35.16%, see Table 22. Bhutan has no F-Measure score for this group above 70% across 1999 to 2011, where the average F-Measure score is 41.75%, see Table 25. The group F-Measure scores for Pakistan range from 28.57% to 85.71% between years 1999 – 2011 with an average F-Measure score of 49.45%, see Table 28. Countries with more occurrences of 1 ratings have lower overall F-Measure scores such as Austria, Bhutan, or Pakistan (see Tables 23, 26, and 29). Another major factor in the reason why ratings of 1 bring the F-Measure lower, especially for this group, is because of the qualitative nature of the coding schemes and country reports. CIRI coders use subjectivity when assigning rating values to human rights practices, which are not well-documented in the CIRI Human Rights Data Project Coding Manual [Cingranelli14B]. As shown in the Tables 21, 24, and 27 below, the more

incorrectly matched Key and Response Set human rights practice ratings, the lower the overall F-Measure score because this group has a higher impact on the evaluation results.

Year	Worker Rights		Freedom of Dom. Movement		Religious Freedom		Freedom of Foreign Movement	
	Key	Response	Key	Response	Key	Response	Key	Response
1999	2	1	2	0	1	2	2	0
2000	1	1	2	0	1	2	2	0
2001	2	0	2	2	1	0	2	2
2002	2	0	2	0	1	0	2	0
2003	2	0	2	0	1	2	2	0
2004	2	1	2	1	1	1	2	1
2005	2	1	2	1	1	1	2	1
2006	2	1	2	1	1	1	2	1
2007	2	1	2	1	1	1	2	1
2008	1	1	2	1	2	1	2	1
2009	2	1	2	0	1	0	2	0
2010	2	1	2	0	2	0	2	0
2011	1	0	2	0	2	2	2	0
Year	Electoral Self-Determination		Freedom of Speech		Freedom of Assembly & Association			
	Key	Response	Key	Response	Key	Response		
1999	2	2	1	2	1	2		
2000	2	2	1	2	1	2		
2001	2	0	1	2	1	2		
2002	2	2	1	0	1	2		
2003	2	2	1	2	1	2		
2004	2	2	1	2	1	2		
2005	2	2	1	2	1	2		
2006	2	2	1	1	2	2		
2007	2	2	1	1	2	2		
2008	2	2	1	1	2	2		
2009	2	2	1	1	2	2		
2010	2	2	1	1	2	2		
2011	2	2	1	0	2	2		

Table 21. Austria Empowerment Rights Key and Response Data

Year	F-Measure	Correct	Missing
1999	0.1429	1	6
2000	0.2857	2	5
2001	0.2857	2	5
2002	0.1429	1	6
2003	0.1429	1	6
2004	0.2857	2	5
2005	0.2857	2	5
2006	0.5714	4	3
2007	0.5714	4	3
2008	0.5714	4	3
2009	0.4286	3	4
2010	0.4286	3	4
2011	0.4286	3	4

Table 22. Austria Empowerment Rights F-Measure Scores

Year	F-Measure	Key				Response			
		0	1	2	3	0	1	2	3
1999	0.2857	0	4	10	0	3	3	8	0
2000	0.2857	1	5	7	1	3	3	8	0
2001	0.3571	0	4	9	1	5	1	8	0
2002	0.2143	0	4	9	1	10	0	4	0
2003	0.2857	0	5	8	1	5	0	9	0
2004	0.3571	0	4	9	1	4	4	6	0
2005	0.4286	0	4	9	1	2	4	8	0
2006	0.5	0	3	10	1	3	5	6	0
2007	0.5	0	3	9	2	2	7	5	0
2008	0.5	0	3	10	1	2	7	5	0
2009	0.4286	0	3	11	0	5	5	4	0
2010	0.5	0	2	10	2	4	5	5	0
2011	0.4286	0	3	9	2	5	2	7	0

Table 23. Austria Overall Evaluation Results

Year	Worker Rights		Freedom of Dom. Movement		Religious Freedom		Freedom of Foreign Movement	
	Key	Response	Key	Response	Key	Response	Key	Response
1999	0	2	1	2	0	0	1	2
2000	0	2	2	0	0	0	2	0
2001	0	2	1	0	0	0	2	0
2002	0	2	2	0	0	0	2	0
2003	0	2	1	2	1	0	1	2
2004	0	2	1	2	0	0	1	2
2005	0	2	1	0	0	0	1	0
2006	0	1	1	0	0	0	1	0
2007	0	0	0	0	1	0	1	0
2008	0	0	0	0	2	0	1	0
2009	0	0	0	0	1	0	1	0
2010	0	2	0	2	1	2	2	2
2011	0	0	0	1	1	2	2	1
Year	Electoral Self-Determination		Freedom of Speech		Freedom of Assembly & Association			
	Key	Response	Key	Response	Key	Response		
1999	0	0	0	0	0	0		
2000	0	0	0	0	0	0		
2001	0	0	0	0	0	0		
2002	0	0	0	0	0	0		
2003	0	0	1	0	0	0		
2004	0	0	1	0	1	0		
2005	0	0	1	2	0	2		
2006	0	0	1	0	0	0		
2007	1	2	1	2	0	0		
2008	0	2	1	2	0	1		
2009	0	2	1	1	0	1		
2010	0	2	1	0	0	0		
2011	1	2	2	2	0	0		

Table 24. Bhutan Empowerment Rights Key and Response Data



Year	F-Measure	Correct	Missing
1999	0.5714	4	3
2000	0.5714	4	3
2001	0.5714	4	3
2002	0.5714	4	3
2003	0.2857	2	5
2004	0.2857	2	5
2005	0.2857	2	5
2006	0.4286	3	4
2007	0.4286	3	4
2008	0.2857	2	5
2009	0.4286	3	4
2010	0.2857	2	5
2011	0.4286	3	4

Table 25. Bhutan Empowerment Rights F-Measure Scores

Year	F-Measure	Key				Response			
		0	1	2	3	0	1	2	3
1999	0.4286	7	4	3	0	6	1	5	2
2000	0.4286	7	3	4	0	10	0	4	0
2001	0.4286	7	2	5	0	11	0	3	0
2002	0.4286	7	2	5	0	9	0	3	2
2003	0.3571	5	6	3	0	6	0	6	2
2004	0.4286	4	5	5	0	5	0	7	2
2005	0.5	5	3	6	0	5	0	7	2
2006	0.5	5	4	5	0	7	3	4	0
2007	0.6429	4	5	4	1	6	2	6	0
2008	0.4286	5	4	4	1	5	4	5	0
2009	0.5714	5	5	3	1	5	2	5	2
2010	0.4286	5	3	5	1	4	1	7	2
2011	0.4286	4	3	7	0	2	4	8	0

Table 26. Bhutan Overall Evaluation Results

Year	Worker Rights		Freedom of Dom. Movement		Religious Freedom		Freedom of Foreign Movement	
	Key	Response	Key	Response	Key	Response	Key	Response
1999	0	0	1	1	0	0	1	1
2000	1	0	1	0	0	0	1	0
2001	0	0	1	2	0	0	1	2
2002	0	0	1	1	0	0	1	1
2003	0	0	1	2	0	0	1	2
2004	0	2	2	0	0	0	2	0
2005	0	0	2	0	0	0	1	0
2006	0	0	2	0	0	0	1	0
2007	0	0	1	0	0	0	0	0
2008	0	0	1	0	0	0	0	0
2009	0	1	1	0	0	0	1	0
2010	0	0	1	0	0	0	1	0
2011	0	0	1	1	0	0	1	1
Year	Electoral Self-Determination		Freedom of Speech		Freedom of Assembly & Association			
	Key	Response	Key	Response	Key	Response		
1999	1	0	1	0	1	0		
2000	0	0	2	0	1	0		
2001	1	0	2	0	0	2		
2002	1	0	1	0	0	0		
2003	0	0	0	0	1	0		
2004	1	0	1	0	1	1		
2005	1	0	0	0	1	1		
2006	1	0	1	0	1	0		
2007	1	0	1	0	0	0		
2008	0	0	0	0	0	2		
2009	0	1	0	0	0	2		
2010	0	1	0	0	1	2		
2011	0	0	0	0	1	2		

Table 27. Pakistan Empowerment Rights Key and Response Data

Year	F-Measure	Correct	Missing
1999	0.5714	4	3
2000	0.2857	2	5
2001	0.2857	2	5
2002	0.7143	5	2
2003	0.5714	4	3
2004	0.2857	2	5
2005	0.5714	4	3
2006	0.2857	2	5
2007	0.5714	4	3
2008	0.7143	5	2
2009	0.2857	2	5
2010	0.4286	3	4
2011	0.8571	6	1

Table 28. Pakistan Empowerment Rights F-Measure Scores

Year	F-Measure	Key				Response			
		0	1	2	3	0	1	2	3
1999	0.5714	6	7	1	0	11	3	0	0
2000	0.5	7	6	1	0	11	1	2	0
2001	0.5714	8	4	2	0	8	1	5	0
2002	0.7143	8	4	2	0	11	3	0	0
2003	0.6429	9	4	1	0	10	0	4	0
2004	0.4286	7	6	1	0	9	2	3	0
2005	0.5714	8	4	2	0	8	3	3	0
2006	0.4286	8	4	2	0	12	1	1	0
2007	0.5714	10	3	1	0	10	1	3	0
2008	0.7143	12	1	1	0	10	1	3	0
2009	0.5	11	2	1	0	8	3	3	0
2010	0.6429	9	4	1	0	10	1	3	0
2011	0.7143	9	4	1	0	8	2	2	2

Table 29. Pakistan Overall Evaluation Results

### 6.5.5 Women's Rights

There are two human rights practice ratings contained in this group, Women's Economical Rights and Women's Political Rights. It is the group that has the third most effect when evaluating the automated ratings with 14.28% or two out of fourteen human practices. The human rights practice ratings in this group are the most unique in that not only are the coding schemes qualitative, but the rating range is between 0-3. This means that rating values for 1 and 2 become more complex, as it breaks the coding scheme used by most of the human rights practices. A rating of 3 in this group is the highest value and is equal to a 2 rating when compared to other human rights practices. The trends seen in this group is the same as those found in the other groups in that the more occurrences of 1 or 2 ratings, the less accurate the F-Measure score becomes. The trend in this group can be seen in Tables 30 and 32 below where details are broken down.

Year	Women's Economic Rights		Women's Political Rights	
	Key	Response	Key	Response
1999	2	3	2	3
2000	2	0	2	0
2001	2	0	2	0
2002	2	0	2	0
2003	3	2	2	2
2004	3	0	2	0
2005	3	0	2	0
2006	3	2	2	2
2007	3	2	2	2
2008	2	0	3	0
2009	3	3	3	3
2010	3	3	3	3
2011	3	1	2	1

Table 30. Australia Women's Rights Key and Response Data

Year	F-Measure	Correct	Missing
1999	0	0	2
2000	0	0	2
2001	0	0	2
2002	0	0	2
2003	0.5	1	1
2004	0	0	2
2005	0	0	2
2006	0.5	1	1
2007	0.5	1	1
2008	0	0	2
2009	1	2	0
2010	1	2	0
2011	0	0	2

Table 31. Australia Women's Rights F-Measure Scores

Year	Women's Economic Rights		Women's Political Rights	
	Key	Response	Key	Response
1999	1	3	2	3
2000	1	0	2	0
2001	1	0	2	0
2002	1	3	2	3
2003	1	0	1	0
2004	1	1	1	1
2005	1	2	1	2
2006	1	0	1	0
2007	1	0	1	0
2008	1	0	1	0
2009	1	1	1	1
2010	1	2	2	2
2011	0	1	1	1

Table 32. Egypt Women's Rights Key and Response Data

Year	F-Measure	Correct	Missing
1999	0	0	2
2000	0	0	2
2001	0	0	2
2002	0	0	2
2003	0	0	2
2004	1	2	0
2005	0	0	2
2006	0	0	2
2007	0	0	2
2008	0	0	2
2009	1	2	0
2010	0.5	1	1
2011	0.5	1	1

Table 33. Egypt Women's Rights F-Measure Scores

#### 6.5.6 Results Summary

The CIRI dataset is comprised of 71.42% of qualitative human rights practice indicators and 28.57% of quantitative human rights practice indicators. As mentioned above, the Independent Judiciary group is the least impactful group with only one human practice and the most impactful group is Empowerment Rights which contains seven human rights practice indicators. The Physical Integrity group is the most accurate because of the quantitative coding schemes from the CIRI Coding Manual [Cingranelli14B]. The Women's Rights group F-Measure score was the most incorrect of the four groups. The average F-Measure score range for this group ranged between 11.53% and 42.30%. The group that had the most variance was Independent Judiciary because this group had one human rights practice indicator so the F-Measure could be either 0% or 100%. Physical Integrity average F-Measure score range was between 50% and 88.46%. The Empowerment Rights group average F-Measure score was between 35.16% and 75.82%.

The average F-Measure scores for all sample countries were between 39.01% and 74.17%, see Appendix C. Overall, the evaluation results show that the CIRI Automated System did better at rating Physical Integrity human rights practice indicators with more correctness, but rated Empowerment Rights human rights practice indicators with more consistency and less variance, but less accuracy.

## Chapter 7

### CONCLUSION

The objective of this thesis was to create an automated process for generating CIRI Human Rights Practice ratings based on the ratings provided by the now discontinued CIRI Human Rights Data Project to continue the work this group had started. The ratings in the CIRI dataset were manually annotated by the CIRI coders based on criteria outlined in the CIRI documentation. The reason why the project was discontinued was not given, but the criteria followed by the CIRI coders in the CIRI documentation provided a solid basis as to how to create the new automated process, the CIRI Automated System. Since the evaluation of years 1999 to 2011 involved parsing the text of the country reports from the United State Department of State, the natural processing language software, GATE, seemed to best fit the needs of the research.

The availability of open source software made developing customized processes for annotation, generation, and evaluation of the country report documents and CIRI Human Rights practice ratings easier. The open source database MySQL was used to store the source and generated CIRI Ratings to be used for evaluation purposes. The open source HTML parsers helped to easier retrieve and remove unneeded HTML markup from the United States Department of State Human Rights website. Lastly, GATE and the free version of IBM's Sentiment analysis software provided tools to create a design that



analyzed text from the country reports and assigning a rating to that text using the coding schemes from the CIRI documentation.

Based on the results gathered in the evaluation portion of this thesis, it is clear that the CIRI Automated System is better at detecting quantitative text and assigning the appropriate rating compared to doing the same tasks for qualitative text. Therefore, in order for an automated process to be created based on the CIRI manual process more strict rules need to be implemented so that an acceptable ratings baseline can be found when the CIRI ratings are automatically generated. This research has shown, however, that an automated process can be created to continue assigning ratings to the country reports that the CIRI project stopped evaluating in 2014. In addition to an automated system, there needs to be manual intervention when deciding on the list of patterns to use and also to verify if the generated ratings are acceptable.

The CIRI Automated System can be used to start the discussion of using an automated process to rate or assign values to countries regarding human rights practices. The use of open source software and natural language processing software such as GATE can be valuable and the CIRI Automated System can be used as the foundation. One limitation to this research was accounting for the difference between the standard precision and recall formulas and GATE's own variation that deals with Annotations and feature values. Another limitation was creating the necessary patterns that are used when detecting the appropriate context of text. Future work includes improving the CIRI Automated System's qualitative pattern recognition. Different natural language processing techniques such as

machine learning with Apache's OpenNLP software or the fully featured machine learning and text classification services of IBM's Watson Natural Language Understanding could give better results.

## REFERENCES

### Print Publications:

[Alence15]

Alence, R., "Mining For Meaning: What the APRM Says (and Doesn't Say) about Africa's Extractive Industries", Electoral Institute for Sustainable Democracy in Africa, M. Roberts, ed., EISA, South Africa, 2015.

[Cunningham02]

Cunningham, H., D. Maynard, K. Bontcheva, and V. Tablan, "GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications", Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL), (July, 2002), pp. 168-175.

[Cunningham13]

Cunningham, H., V. Tablan, A. Roberts, and K. Bontcheva, "Getting More Out of Biomedical Documents with GATE's Full Lifecycle Open Source Text Analytics", PLoS Computational Biology 9, 2 (February, 2013), doi:10.1371/journal.pcbi.1002854.

[Hevner04]

Hevner, A. R., S. T. March, J. Park, and S. Ram, "Design science in information systems research", MIS Quarterly 28, 1 (March, 2004), pp. 75-105.

[Katakis08]

Katakis, I., G. Tsoumakas, and I. Vlahavas, "Multilabel Text Classification for Automated Tag Suggestion", Antwerp: International Workshop at The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, (2008), pp. 75-83.

[Liu12]

Liu, B., "Sentiment Analysis and Opinion Mining", Synthesis lectures on human language technologies 5, 1 (May, 2012), pp. 1-167.

[Maynard12]

Maynard, D. and M. A. Greenwood, "Large Scale Semantic Annotation, Indexing, and Search at The National Archives", In Proceedings of LREC 2012, (May, 2012).

[Minhas15]

Minhas, S., J. Ulfelder, and M. D. Ward, "Mining texts to efficiently generate global data on political regime types", Research and Politics 2, 3 (August, 2015), doi:10.1177/2053168015589217.

[Nilesh09]

Nilesh, D., et al., "A Web of Concepts", Proceedings of the twenty-eighth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, (June, 2009), pp. 1-12.

[Witten04]

Witten, I. H., K. J. Don, M. Dewship, and V. Tablan, "Text mining in a digital library", International Journal on Digital Libraries 4, 1 (2004), pp. 56-59.

Electronic Sources:

[Amnesty International16]

Amnesty International, <https://www.amnesty.org/en/who-we-are/>, 2016, last accessed March 20, 2018

[Cingranelli14A]

Cingranelli, D. L., D. L. Richards, and K. C. Clay, "CIRI Human Rights Project", <http://www.humanrightsdata.com/p/data-documentation.html>, last revision May 2014, last accessed April 30, 2018

[Cingranelli14B]

Cingranelli, D. L., D. L. Richards, and K. C. Clay, "The Cingranelli-Richards (CIRI) Human Rights Data Project Coding Manual Version 5.20.14", <http://www.humanrightsdata.com/p/data-documentation.html>, last revision May 20, 2014, last accessed May 1, 2018

[Cingranelli14C]

Cingranelli, D. L., D. L. Richards, and K. C. Clay, "The CIRI Human Rights Dataset", <http://www.humanrightsdata.com/p/data-documentation.html>, last revision April 14, 2014, last accessed May 1, 2018

[Cunningham14]

Cunningham, H., et al., "Developing Language Processing Components with GATE Version 8", University of Sheffield Department of Computer Science, <https://gate.ac.uk/releases/gate-8.2-build5482-ALL/doc/tao/split.html>, November 17, 2014, last accessed May 2, 2018

[Farrell16]

Farrell, J. A., "Anatomy of a Compiler", <http://www.cs.man.ac.uk/~pjj/farrell/compmain.html>, October 2016, last accessed March 20, 2018

[Hearst13]

Hearst, M., "What is Text Mining?", <http://people.ischool.berkeley.edu/~hearst/text-mining.html>, October 17, 2013, last accessed March 20, 2018

[Hedley17]

Hedley, J., Jsoup, <https://jsoup.org>, last revision 2017, last accessed March 20, 2018

[Hotho05]

Hotho, A., A. Nürnberger, and G. Paaß, "A Brief Survey of Text Mining", <https://pdfs.semanticscholar.org/9643/0cc91ed91fd2d4042fa6fcb7ecf4005d77a7.pdf>, May 13, 2005, last accessed March 20, 2018

[IBM17]

IBM. "Natural Language Understanding", <https://www.ibm.com/watson/developercloud/doc/natural-language-understanding/>, 2017, last accessed March 20, 2018

[Mekala08]

Swapna, M., "Social network analysis application using GATE", available only on CD-ROM by request from UNF, <https://www.unf.edu/contact/>

[Techopedia16]

Techopedia. "What is Named Entity Recognition?", <https://www.techopedia.com/definition/13825/named-entity-recognition-ner>, October 2016, last accessed March 20, 2018

[The University of Sheffield16]

The University of Sheffield. GATE Overview, <https://gate.ac.uk/overview.html>, October 2016, last accessed March 20, 2018

[United States Department of State16]

United States Department of State, Human Rights Reports, <http://www.state.gov/j/drl/rls/hrrpt/>, 2016, last accessed February 25, 2018

[Yang08]

Yang, Y., and T. Joachims, "Text Categorization", [http://www.scholarpedia.org/article/Text\\_categorization](http://www.scholarpedia.org/article/Text_categorization), 2008, last accessed March 20, 2018

## APPENDIX A

### Gazetteer list – Report Sections

- Arbitrary or Unlawful Deprivation of Life
- Disappearance
- Torture and Other Cruel, Inhuman, or Degrading Treatment or Punishment
- Prison and Detention Center Conditions
- Arbitrary Arrest or Detention
- Role of the Police and Security Apparatus
- Arrest Procedures and Treatment While in Detention
- Denial of Fair Public Trial
- Trial Procedures
- Political Prisoners and Detainees
- Regional Human Rights Court Decisions
- Civil Judicial Procedures and Remedies
- Arbitrary Interference with Privacy, Family, Home, or Correspondence
- Freedom of Speech and Press
- Status of Freedom of Speech and Press
- Internet Freedom
- Academic Freedom and Cultural Events
- Freedom of Peaceful Assembly and Association
- Freedom of Religion

- Freedom of Movement, Internally Displaced Persons, Protection of Refugees, and Stateless Persons
- Elections and Political Participation
- Women
- Children
- Anti-Semitism
- Trafficking in Persons
- Persons with Disabilities
- National/Racial/Ethnic Minorities
- Indigenous People
- Societal Abuses, Discrimination, and Acts of Violence Based on Sexual Orientation and Gender Identity
- Other Societal Violence or Discrimination
- Freedom of Association and the Right to Collective Bargaining
- Prohibition of Forced or Compulsory Labor
- Prohibition of Child Labor and Minimum Age for Employment
- Acceptable Conditions of Work
- Executive Summary

## APPENDIX B

```

Phase: NumberLetter
Input: Token Money Date Number Lookup Person

Options: control = appelt

Rule: LookupMatch

Priority: 100

(((Token.category == "NNP"):nnpLabel|

({Person} {Token.category == "NNS"}:personLabel|

({Token.kind == "word", Token.orth == "allCaps"} {Token.kind == "number", Token.category ==
"CD"}):pageNumber|({Token.category == "CD", Token.length ==
4}|{Money}|{Date}|({Date} {Token.orth == "allCaps"})):dateMoney|(((Token.string =~
"[Ss]ection"))*{Token.category == "CD"} {Token.kind == "punctuation"} {Token.category ==
"SYM"}))|

({Token.string =~ "[Ss]ection"} {Token.kind == "number"})):sectionNumber|

({Number}):numberLabel| ({Lookup.majorType == "year"}):yearOnly({Number}
{Lookup.majorType == "date_unit"}):dateUnit

):match

-->

:nnpLabel.Exclude = {rule = "nnpRule"},

:personLabel.Exclude = {rule = "personName"},

:pageNumber.Exclude = {rule = "pageNumber"},

:dateMoney.Exclude = {rule = "dateMoney"},

:numberLabel.Exclude = {rule = "numberRule"},

:match.Exclude = {rule = "allMatch"},

:dateUnit.Exclude = {rule = "dateUnit"},

:yearOnly.Exclude = {rule = "yearOnly"}  (Ctrl)

```

Figure 20. NumberLetter



```

Imports: {

import static gate.Utils.*;

}

Phase: HtmlDelimiter

Input: Lookup Token Date Sentence

Options: control = appelt

Rule: StartDelim

Priority: 100

({Sentence} {Sentence} {Date.ruleFinal == "YearOnlyFinal"} {Date.ruleFinal ==
"DateOnlyFinal"}):startMatch

-->

:startMatch.Separator = {rule = "delimiter"}

Rule: EndDelim

Priority: 100

({Token.root == "["} ({Token.kind == "word"})[3]{Token.root == "]" }):endMatch

-->

:endMatch.Separator = {rule = "delimiter"}|

```

Figure 21. HTMLDelimiter

```

Phase: WomenPoliticalRatio

Input: Token SpaceToken Sentence IncludeNumber Gender

Options: control = appelt

Rule: PoliticalRatio

(
  ({IncludeNumber}):firstNumber{SpaceToken} {Token.string == "of"}
  {SpaceToken}({IncludeNumber}):secondNumber|
  ({IncludeNumber}):firstNumber2{SpaceToken} {Gender} {SpaceToken} {Token.string == "in"}
  {SpaceToken} {Token.string == "the"}{SpaceToken}({IncludeNumber}):secondNumber2
):calTag
-->

:firstNumber.WomenPoliticalRatio = {position = "1"},
:secondNumber.WomenPoliticalRatio = {position = "2"},
:firstNumber2.WomenPoliticalRatio = {position = "1"},
:secondNumber2.WomenPoliticalRatio = {position = "2"},
:calTag.WomenPoliticalRatio = {kind = "ratio"}

```

Figure 22. WomenPolitialRatio

```

Phase: StartSentence
Input: Token Sentence Ciri_Header
Options: control = appelt

Rule: Header
Priority: 100
(
{Ciri_Header contains {Token.string =~ "(?i)RESPECT FOR HUMAN RIGHTS"}}
):header
-->
:header.Ciri_Header = {majorType = "ciri", minorType = "Header"}|

```

Figure 23. CIRIHeaderPatterns (1)

```

Rule: EndDocument
Priority: 100
(
{Sentence contains {Token.string == "["}, Sentence contains {Token.string == "]"}}
):enddocument
-->
:enddocument.Ciri_Header = {majorType = "ciri", minorType = "EndDocument"}|

```

Figure 24. CIRIHeaderPatterns (2)

```

Rule: StartingSentence
Priority: 100
({Sentence}):firstSentence
-->
{
AnnotationSet tokenAS = (gate.AnnotationSet) bindings.get("firstSentence");
AnnotationSet tokens = gate.Utills.getContainedAnnotations(inputAS, tokenAS, "Token");
List<Annotation> sentList = gate.Utills.inDocumentOrder(tokens);
Annotation first = sentList.get(0);
Long start = first.getStartNode().getOffset();

if(start == 0 || start == 1){
FeatureMap features = Factory.newFeatureMap();
features.put("majorType", "ciri");
features.put("minorType", "executive_summary");
gate.Utills.addAnn(outputAS, first, "Ciri_Level0", features);
}
}

```

Figure 25. CIRIHeaderPatterns (3)

## APPENDIX C

Country	Overall F-Measure	Independent Judiciary F-Measure	Physical Integrity F-Measure	Empowerment Rights F-Measure	Women's Rights F-Measure
Angola	49.99%	100%	50%	48.35%	30.76%
Sudan	65.93%	7.69%	78.57%	74.72%	26.92%
Australia	74.17%	100%	88.46%	75.82%	26.92%
Cambodia	53.29%	100%	51.92%	56.04%	23.07%
Austria	39.01%	100%	42.30%	35.16%	15.38%
Denmark	69.23%	100%	88.46%	67.03%	23.07%
Bahrain	53.29%	7.70%	63.46%	61.65%	19.23%
Egypt	43.40%	16.38%	53.84%	47.25%	23.07%
Bhutan	46.15%	46.15%	71.15%	41.75%	11.53%
Pakistan	58.24%	38.46%	86.53%	49.45%	42.30%
Canada	59.34%	100%	59.61%	64.83%	19.23%
Colombia	45.60%	0%	88.46%	35.16%	23.07%

Table 34. Overall Sample Country Evaluations

### 11.1 Africa

Year	F-Measure
1999	0.4286
2000	0.6429
2001	0.5714
2002	0.7143
2003	0.5
2004	0.4286
2005	0.3571
2006	0.5714
2007	0.3571
2008	0.5714
2009	0.5714
2010	0.5
2011	0.2857

Table 35. Angola

<b>Year</b>	<b>F-Measure</b>
1999	0.7143
2000	0.8571
2001	0.7857
2002	0.7857
2003	0.7143
2004	0.5714
2005	0.5714
2006	0.6429
2007	0.6429
2008	0.7143
2009	0.5714
2010	0.4286
2011	0.5714

Table 36. Sudan

## 11.2 East Asia and the Pacific

<b>Year</b>	<b>F-Measure</b>
1999	0.7857
2000	0.5
2001	0.5714
2002	0.5
2003	0.7143
2004	0.7857
2005	0.7857
2006	0.8571
2007	0.8571
2008	0.7143
2009	0.8571
2010	0.9286
2011	0.7857

Table 37. Australia

<b>Year</b>	<b>F-Measure</b>
1999	0.7143
2000	0.3571
2001	0.5714
2002	0.2857
2003	0.4286
2004	0.5714
2005	0.6429
2006	0.5
2007	0.5
2008	0.5714
2009	0.7143
2010	0.5714
2011	0.5

Table 38. Cambodia

### 11.3 Europe and Eurasia

<b>Year</b>	<b>F-Measure</b>
1999	0.2857
2000	0.2857
2001	0.3571
2002	0.2143
2003	0.2857
2004	0.3571
2005	0.4286
2006	0.5
2007	0.5
2008	0.5
2009	0.4286
2010	0.5
2011	0.4286

Table 39. Austria

Year	F-Measure
1999	0.8571
2000	0.7857
2001	0.7143
2002	0.8571
2003	0.7857
2004	0.7143
2005	0.6429
2006	0.6429
2007	0.5
2008	0.5714
2009	0.5714
2010	0.5714
2011	0.7857

Table 40. Denmark

#### 11.4 Near East and North Africa

Year	F-Measure
1999	0.5714
2000	0.5714
2001	0.5714
2002	0.3571
2003	0.5
2004	0.5
2005	0.5
2006	0.5
2007	0.4286
2008	0.5714
2009	0.6429
2010	0.5714
2011	0.6429

Table 41. Bahrain



Year	F-Measure
1999	0.2143
2000	0.5714
2001	0.2857
2002	0.2143
2003	0.2857
2004	0.5
2005	0.5
2006	0.5
2007	0.2857
2008	0.4286
2009	0.7143
2010	0.6429
2011	0.5

Table 42. Egypt

## 11.5 South and Central Asia

Year	F-Measure
1999	0.4286
2000	0.4286
2001	0.4286
2002	0.4286
2003	0.3571
2004	0.4286
2005	0.5
2006	0.5
2007	0.6429
2008	0.4286
2009	0.5714
2010	0.4286
2011	0.4286

Table 43. Bhutan

Year	F-Measure
1999	0.5714
2000	0.5
2001	0.5714
2002	0.7143
2003	0.6429
2004	0.4286
2005	0.5714
2006	0.4286
2007	0.5714
2008	0.7143
2009	0.5
2010	0.6429
2011	0.7143

Table 44. Pakistan

## 11.6 Western Hemisphere

Year	F-Measure
1999	0.7143
2000	0.7143
2001	0.6429
2002	0.7143
2003	0.7857
2004	0.4286
2005	0.5714
2006	0.5714
2007	0.5
2008	0.5
2009	0.5
2010	0.4286
2011	0.6429

Table 45. Canada

Year	F-Measure
1999	0.5
2000	0.5
2001	0.2857
2002	0.4286
2003	0.4286
2004	0.5714
2005	0.5714
2006	0.6429
2007	0.3571
2008	0.4286
2009	0.5
2010	0.2857
2011	0.4286

Table 46. Colombia

## VITA

Joshua Joiner has Bachelor of Science degree from the University of North Florida in Information Systems and expects to receive a Master of Science in Information Systems in spring 2018. Joshua is currently employed as a Web Applications Developer at Black Knight, Inc. and has been full time in this position for 5 years.

Joshua specializes in front end and back end web development and is interested in creating user friendly and valuable web applications. Joshua has professional experience developing within a Java-based technology stack using web frameworks such as Spring or Struts2 and other standard web technologies such as JavaScript and CSS.