REES: A Large-Scale Relation and Event Extraction System

Chinatsu Aone
SRA International, Inc.
4300 Fair Lakes Court
Fairfax, VA 22033
aonec@verdi.sra.com

Mila Ramos-Santacruz
SRA International, Inc.
4300 Fair Lakes Court
 Fairfax, VA 22033
mila@verdi.sra.com

Abstract
This paper reports on a large-scale, end-to-end relation and event extraction system. At present, the system extracts a total of 100 types of relations and events, which represents a much wider coverage than is typical of extraction systems. The system consists of three specialized pattern-based tagging modules, a high-precision coreference resolution module, and a configurable template generation module. We report quantitative evaluation results, analyze the results in detail, and discuss future directions.

Introduction
One major goal of information extraction (IE) technology is to help users quickly identify a variety of relations and events and their key players in a large volume of documents. In contrast with this goal, state-of-the-art information extraction systems, as shown in the various Message Understanding Conferences (MUCs), extract a small number of relations and events. For instance, the most recent MUC, MUC-7, called for the extraction of 3 relations (person-employer, maker-product, and organization-location) and 1 event (spacecraft launches). Our goal is to develop an IE system which scales up to extract as many types of relations and events as possible with a minimum amount of porting effort combined with high accuracy. Currently, REES handles 100 types of relations and events, and it does so in a modular, configurable, and scalable manner.

Below, Section 1 presents the ontologies of relations and events that we have developed.

Table: Relations

| Place Relations             | Artifact Relations       |
|-----------------------------|--------------------------|
| Place-Name&Aliases          | Artifact-Name&Aliases    |
| Place-Type                  | Artifact-Type            |
| Place-Subtype               | Artifact-Subtype         |
| Place-Descriptor            | Artifact-Descriptor      |
| Place-Country               | Artifact-Maker           |
|                            | Artifact-Owner           |
| Organization Relations      | Person Relations         |
| Org-Name&Aliases            | Person-Name&Aliases      |
| Org-Descriptor              | Person-Type              |
| Org-FoundationDate          | Person-Subtype           |
| Org-Nationality             | Person-Descriptor        |
| Org-TickerSymbol            | Person-Honorific         |
| Org-Location                | Person-Age               |
| Org-ParentOrg               | Person-PhoneNumber        |
| Org-Owner                   | Person-Nationality       |
| Org-Founder                 | Person-Affiliation       |
| Org-StockMarket             | Person-Sibling           |
|                            | Person-Spouse            |
|                            | Person-Parent            |
|                            | Person-Grandparent       |
"Events" are extracted along with their event participants, e.g., "who did what to whom when and where?" For example, for a BUYING event, REES extracts the buyer, the artifact, the seller, and the time and location of the BUYING event. REES currently covers 61 types of events, as shown below.

**Table 1: Relation Ontology**

| Events          | Transaction          |
|-----------------|----------------------|
| Vehicle         | Transaction          |
| Vehicle departs | Buy artifact         |
| Vehicle arrives | Sell artifact        |
| Spacecraft launch| Import artifact      |
| Vehicle crash   | Export artifact      |
|                 | Give money           |
| Personnel Change| Business             |
| Hire            | Start business       |
| Terminate contract | Close business     |
| Promote         | Make artifact        |
| Succeed         | Acquire company      |
| Start office    | Sell company         |
|                 | Sue organization     |
|                 | Merge company        |
| Crime           | Financial            |
| Sexual assault  | Currency moves up    |
| Steal money     | Currency moves down  |
| Seize drug      | Stock moves up       |
| Indict          | Stock moves down     |
| Arrest          | Stock market moves up|
| Try             | Stock market moves down|
| Convict         | Stock index moves up |
| Sentence        | Stock index moves down|
| Jail            |                      |

**Table 2: Event Ontology**

Figures 1 and 2 show sample relation and event templates. Figure 1 shows a Person-Affiliation relation template for "Frank Ashley, a spokesman for Occidental Petroleum Corp."

```
<PERSON_AFFILIATION-AP8802230207-54> :=
  TYPE: PERSON_AFFILIATION
  PERSON: [TE for "Frank Ashley"]
  ORG: [TE for "Occidental Petroleum"]
```

**Figure 1: Example of Relation Template**

Figure 2 shows an Attack Target event template for the sentence "an Iraqi warplane attacked the frigate Stark with missiles May 17, 1987."

```
<ATTACK_TARGET-AP8804160078-12> :=
  TYPE: CONFLICT
  SUBTYPE: ATTACK_TARGET
  Attacker: [TE for "an Iraqi warplane"]
  TARGET: [TE for "the frigate Stark"]
  WEAPON: [TE for "missiles"]
  TIME: "May 17, 1987"
  PLACE: [TE for "the gulf"]
  COMMENT: "attacked"
```

**Figure 2: Example of Event Template**

## 2 System Architecture and Components

Figure 3 illustrates the REES system architecture. REES consists of three main components: a tagging component (cf. Section 2.1), a co-reference resolution module (cf. Section 2.2), and a template generation module (cf. Section 2.3). Figure 3 also illustrates that the user may run REES from a Graphical User Interface (GUI) called TemplateTool (cf. Section 2.4).

### 2.1 Tagging Modules

The tagging component consists of three modules as shown in Figure 3: NameTagger, NPTagger and EventTagger. Each module relies on the same pattern-based extraction engine, but uses different sets of patterns. The NameTagger recognizes names of people, organizations, places, and artifacts (currently only vehicles).
The NPTagger then takes the XML-tagged output of the NameTagger through two phases. First, it recognizes non-recursive Base Noun Phrase (BNP) (our specifications for BNP resemble those in Ramshaw and Marcus 1995). Second, it recognizes complex NPs for only the four main semantic types of NPs, i.e., Person, Organization, Location, and Artifact (vehicle, drug and weapon). It makes post-modifier attachment decisions only for those NPs that are crucial to the extraction at hand. During this second phase, relations which can be recognized locally (e.g., Age, Affiliation, Maker) are also recognized and stored using the XML attributes for the NPs. For instance, the XML tag for “President of XYZ Corp.” below holds an AFFILIATION attribute with the ID for “XYZ Corp.”

```
<PNP ID="03" AFFILIATION="04">President of <ENTITY ID="04">XYZ Corp.</ENTITY></PNP>
```

Building upon the XML output of the NPTagger, the EventTagger recognizes events applying its lexicon-driven, syntactically-based generic patterns. These patterns tag events in the presence of at least one of the arguments specified in the lexical entry for a predicate. Subsequent patterns try to find additional arguments as well as place and time adjunct information for the tagged event. As an example of the EventTagger’s generic patterns, consider the simplified pattern below. This pattern matches on an event-denoting verb that requires a direct object of type weapon (e.g., “fire a gun”)

```
& {AND $VP {ARG2_SYN=DO} {ARG2_SEM=WEAPON}}
{AND $ARTIFACT {SUBTYPE=WEAPON}}
```

The important aspect of REES is its declarative, lexicon-driven approach. This approach requires a lexicon entry for each event-denoting word, which is generally a

---

1 $\&=concatenation, \text{AND}=\text{Boolean operator}, \$VP \text{ and } \$ARTIFACT \text{ are macro references for complex phrases.}
verb. The lexicon entry specifies the syntactic and semantic restrictions on the verb's arguments. For instance, the following lexicon entry is for the verb “attack.” It indicates that the verb “attack” belongs to the CONFLICT ontology and to the ATTACK_TARGET type. The first argument for the verb “attack” is semantically an organization, location, person, or artifact (ARG1 SEM), and syntactically a subject (ARG1 SYN). The second argument is semantically an organization, location, person or artifact, and syntactically a direct object. The third argument is semantically a weapon and syntactically a prepositional phrase introduced by the preposition “with”.

ATTACK { { {CATEGORY VERB} {ONTOSTY CONFLICT} {TYPE ATTACK_TARGET} {ARG1 SEM {ORGANIZATION LOCATION PERSON ARTIFACT}}} {ARG1 SYN {SUBJECT}} {ARG2 SEM {ORGANIZATION LOCATION PERSON ARTIFACT}} {ARG2 SYN {DO}} {ARG3 SEM{WEAPON}} {ARG3 SYN {WITH}}} }

About 50 generic event extraction patterns, supported by lexical information as shown above, allow extraction of events and their arguments in cases like:

An Iraqi warplane attacked the frigate Stark with missiles May 17, 1987.

This generic, lexicon-driven event extraction approach makes REES easily portable because new types of events can be extracted by just adding new verb entries to the lexicon. No new patterns are required. Moreover, this approach allows for easy customization capability: a person with no knowledge of the pattern language would be able to configure the system to extract new events.

While the tagging component is similar to other pattern-based IE systems (e.g., Appelt et al. 1995; Aone et al. 1998, Yangarber and Grishman 1998), our EventTagger is more portable through a lexicon-driven approach.

2.2 Co-reference Resolution

After the tagging phase, REES sends the XML output through a rule-based co-reference resolution module that resolves:

- definite noun phrases of Organization, Person, and Location types, and
- singular person pronouns: he and she.

Only “high-precision” rules are currently applied to selected types of anaphora. That is, we resolve only those cases of anaphora whose antecedents the module can identify with high confidence. For example, the pronoun rules look for the antecedents only within 3 sentences, and the definite NP rules rely heavily on the head noun matches. Our high-precision approach results from our observation that unless the module is very accurate (above 80% precision), the co-reference module can hurt the overall extraction results by over-merging templates.

2.3 Template Generation Module

A typical template generation module is a hard-coded post-processing module which has to be written for each type of template. By contrast, our Template Generation module is unique as it uses declarative rules to generate and merge templates automatically so as to achieve portability.

2.3.1 Declarative Template Generation

REES outputs the extracted information in the form of either MUC-style templates, as illustrated in Figure 1 and 2, or XML. A crucial part of a portable, scalable system is to be able to output different types of relations and events without changing the template generation code. REES maps XML-tagged output of the co-reference module to templates using declarative template definitions, which specifies the template label (e.g., ATTACK_TARGET), XML attribute names (e.g., ARGUMENT1), corresponding template slot names (e.g., ATTACKER), and the type restrictions on slot values (e.g., string).
2.3.2 Event Merging

One of the challenges of event extraction is to be able to recognize and merge those event descriptions which refer to the same event. The Template Generation module uses a set of declarative, customizable rules to merge co-referring events into a single event. Often, the rules reflect pragmatic knowledge of the world. For example, consider the rule below for the DYING event type. This rule establishes that if two die events have the same subject, then they refer to the same event (i.e., a person cannot die more than once).

\{merge
\{EVENT 1 {AND {SUBTYPE DIE} {PERSON $foo}}\}
\{EVENT 2 {AND {SUBTYPE DIE} {PERSON $foo}}\}\}

2.4 Graphical User Interface (GUI)

For some applications such as database population, the user may want to validate the system output. REES is provided with a Java-based Graphical User Interface that allows the user to run REES and display, delete, or modify the system output. As illustrated in Figure 4, the tool displays the templates on the bottom half of the screen, and the user can choose which template to display. The top half of the screen displays the input document with extracted phrases in different colors. The user can select any slot value, and the tool will highlight the portion of the input text responsible for the slot value. This feature is very useful in efficiently verifying system output. Once the system’s output has been verified, the resulting templates can be saved and used to populate a database.

3 System Evaluation

The table below shows the system’s recall, precision, and F-Measure scores for the training set (200 texts) and the blind set (208 texts) from about a dozen news sources. Each set contains at least 3 examples of each type of relations and events. As we mentioned earlier, “relations” includes MUC-style TEs and TRs.

| Text Set | Task    | Templates in keys | R | P | F-M |
|----------|---------|-------------------|---|---|-----|
| Train    | Rel.    | 9955              | 76| 74| 75.35|
|          | Events  | 2525              | 57| 74| 64.57|
|          | Rel. &  | 10707             | 74| 74| 73.95|
|          | Events  |                   |   |   |      |
| Blind    | Rel.    | 8938              | 74| 74| 73.74|
|          | Events  | 2020              | 42| 75| 53.75|
|          | Rel. &  | 9526              | 69| 74| 71.39|
|          | Events  |                   |   |   |      |

Table 3: Evaluation Results

The blind set F-Measure for 31 types of relations (73.95%) exceeded our initial goal of 70%. While the blind set F-Measure for 61 types of events was 53.75%, it is significant to note that 26 types of events achieved an F-Measure over 70%, and 37 types over 60% (cf. Table 4). For reference, though not exactly comparable, the best-performing MUC-7 system achieved 87% in TE, 76% in TR, and 51% in event extraction.

| F-M in Event types | Event types |
|--------------------|-------------|
| 90-100             | 2 : Buy artifact, Marry |
| 80-89              | 9 : Succeed, Merge company, Kill, Surrender, Arrest, Convict, Sentence, Nominate, Expel. |
| 70-79              | 13 : Die, Sell artifact, Export Artifact, Hire, Start office, Make artifact, Acquire company, Sue organization, Stock Index moves down, Steal money, Indict, Jail, Vehicle crash, Elect, Hold meeting. |

Table 4: Top-performing Event Types
Regarding relation extraction, the difference in the score between the training and blind sets was very small. In fact, the total F-Measure on the blind set is less than 2 points lower than that of the training set. It is also interesting to note that for 8 of the 12 relation types where the F-Measure dropped more than 10 points, the training set includes less than 20 instances. In other words, there seems to be a natural correlation between low number of instances in the training set and low performance in the blind set.

There was a significant drop between the training and blind sets in event extraction: 11 points. We believe that the main reason is that the total number of events in the training set is fairly low: 801 instances of 61 types of events (an average of 13/event), where 35 of the event types had fewer than 10 instances. In fact, 9 out of the 14 event types which scored lower than 40% F-Measure had fewer than 10 examples. In comparison, there were 34,000 instances of 39 types of relations in the training set.

The contribution of the co-reference module is illustrated in the table below. Co-reference resolution consistently improves F-Measures both in training and blind sets. Its impact is larger in relation than event extraction.

| Text set | Task          | Co-reference rules | No co-reference rules |
|----------|---------------|--------------------|-----------------------|
| Training | Relations     | 75.35              | 72.54                 |
|          | Events        | 64.57              | 63.62                 |
|          | Relations & Events | 73.95             | 71.34                 |
| Blind    | Relations     | 73.74              | 72.03                 |
|          | Events        | 53.75              | 53.22                 |
|          | Relations & Events | 71.39             | 69.86                 |

Table 5: Comparative results with and without co-reference rules

In the next two sections, we analyze both false positives and false negatives.
3.1 False Positives (or Precision Errors)

REES produced precision errors in the following cases:

- Most of the errors were due to over-generation of templates. These are mostly cases of co-referring noun phrases that the system failed to resolve. For example: "Panama ... the nation ... this country ... his country"
  
  Rules for the co-reference module are still under development, and at present REES handles only limited types of plural noun phrase anaphora.

- Spurious events resulted from verbs in conditional constructions (e.g., "if ... then...") or from ambiguous predicates. For instance, "appoint" as a POLITICAL event vs. a PERSONNEL_CHANGE event.

- The subject of a verb was misidentified. This is particularly frequent in reduced relative clauses.
  
  Kabul radio said the latest deaths brought to 38 the number of people killed in the three car bomb explosions.
  
  (Wrong subject: "the number of people" as the KILLER instead of the victim)

3.2 False Negatives (or Recall Errors)

Below, we list the most frequent recall errors in the training set.

- Some event arguments are mentioned with event nouns instead of event verbs. The current system does not handle noun-based event extraction.
  
  India's acquisition last month of the nuclear submarine from the Soviet Union...
  
  (SELLER="Soviet Union" and TIME="last month" come with the noun-based event "acquisition."

- Pronouns "it" and "they," which carry little semantic information, are currently not resolved by the co-reference module.
  
  It also has bought three late-1970s vintage Kilo class Soviet submarines and two West German HDW 209 subs
  
  (Missed BUYER=India because of unresolved it.)

- Verb arguments are a conjunction of noun phrases. The current system does not handle coordination of verb arguments.
  
  Hezbollah killed 21 Israelis and 43 of Ladhad's soldiers
  
  (The system gets only the first object: 21 Israelis.)

- Ellipsis cases. The current system does not handle ellipsis.
  
  The two were sentenced to five-year prison terms with hard labor by the state security court...
  
  (Missed PERSON_SENTENCED fill because of unresolved the two.)

- The subject of the event is relatively far from the event-denoting verb:
  
  Vladislav Listev, 38, who brought television interview shows in the style of Phil Donahue or Larry King to Russian viewers and pioneered hard-hitting television journalism in the 1980s, was shot in the heart by unknown assailants and died immediately...
  
  (The system missed subject Vladislav Listev for attack event shot)

- Missed ORG_LOCATION relations for locations that are part of the organization's name.
  
  Larnaca General Hospital
  
  (Missed ORG_LOCATION TR for this and Larnaca.)

We asked a person who is not involved in the development of REES to review the event extraction output for the blind set. This person reported that:

- In 35% of the cases where the REES system completely missed an event, it was because the lexicon was missing the predicate. REES's event predicate lexicon is rather small at present (a total of 140 verbs for 61 event types) and is mostly based on the examples found in the training set.

- In 30% of the cases, the subject or object was elliptical. The system does not currently handle ellipsis.
• In 25% of the cases, syntactic/semantic argument structures were missing from existing lexical entries. It is quite encouraging that simply adding additional predicates and predicate argument structures to the lexicon could significantly increase the blind set performance.

4 Future Directions

We believe that improving co-reference resolution and adding noun-based event extraction capability are critical to achieving our ultimate goal of at least 80% F-Measure for relations and 70% for events.

4.1 Co-reference Resolution

As discussed in Section 3.1 and 3.2, accurate co-reference resolution is crucial to improving the accuracy of extraction, both in terms of recall and precision. In particular, we identified two types of high-payoff co-reference resolution:

• definite noun phrase resolution, especially plural noun phrases
• 3rd person neutral pronouns “it” and “they.”

4.2 Noun-based Event Extraction

REES currently handles only verb-based events. Noun-based event extraction adds more complexity because:

• Nouns are often used in a generic, non-referential manner (e.g., “We see a merger as being in the consumer’s interest”), and

• When referential, nouns often refer to verb-based events, thus requiring noun-verb co-reference resolution (“An F-14 crashed shortly after takeoff... The crash”).

However, noun-based events are crucial because they often introduce additional key information, as the underlined phrases below indicate: While Bush’s meetings with prominent anti-apartheid leaders such as Archbishop Desmond Tutu and Albertina Sisulu are important...

We plan to develop a generic set of patterns for noun-based event extraction to complement the set of generic verb-based extraction patterns.

5 Conclusions

In this paper, we reported on a fast, portable, large-scale event and relation extraction system REES. To the best of our knowledge, this is the first attempt to develop an IE system which can extract such a wide range of relations and events with high accuracy. It performs particularly well on relation extraction, and it achieves 70% or higher F-Measure for 26 types of events already. In addition, the design of REES is highly portable for future addition of new relations and events.

Acknowledgements

This project would have not been possible without the contributions of Arcel Castillo, Lauren Halverson, and Sandy Shinn. Our thanks also to Brandon Kennedy, who prepared the hand-tagged data.

References

Aone, Chinatsu, Lauren Halverson, Tom Hampton, and Mila Ramos-Santacruz. 1998. “SRA: Description of the IE2 System Used for MUC-7.” In Proceedings of the 7th Message Understanding Conference (MUC-7).

Appelt, Douglas E., Jerry R Hobbs, John Bear, David Israel, Megumi Kameyama, Andy Kehler, David Martin, Karen Myers, and Mabry Tyson. 1995. “SRI International FASTUS System: MUC-6 Test Results and Analysis.” In Proceedings of the 6th Message Understanding Conference (MUC-6).

Ramshaw, Lance A., and Mitchell P. Marcus. 1995. “Text Chunking Using Transformation-Based Learning”. In Proceedings of the 3rd ACL Workshop on Very Large Corpora (WVLC95).

Yangarber, Roman and Ralph Grishman. 1998. “NYU: Description of the Proteus/PET System as Used for MUC-7 ST.” In Proceedings of the 6th Message Understanding Conference (MUC-7).