

FASTUS: Extracting Information from Natural-Language Texts

Presented by Warren Wong

Overview

- Motivations for FASTUS
- General Architecture
- Example run through
- Extensibility

Motivations

Want to Extract Information from Text

Context Free System

- Slow
- No need to understand all constructs

Example:

TACITUS

- Text Understanding System
- 36 hours to process 100 messages

Motivations

Finite State System

- Fast (12 minutes for 100 messages)
- Less processing than TACITUS but more *relevant* processing
- Uses less memory in general

FASTUS (Finite State Automaton Text Understanding System)

FASTUS System Architecture

Cascade Finite Automata

5 Steps:

- 1) Complex Words
- 2) Basic Phrases
- 3) Complex Phrases
- 4) Domain Events
- 5) Merging Structures

Example Run Through

Bridgestone Sports Co. said Friday it has set up a joint venture in Taiwan with a local concern and a Japanese trading house to produce golf clubs to be shipped to Japan.

The joint venture, Bridgestone Sports Taiwan Co., capitalized at 20 million new Taiwan dollars, will start production in January 1990 with production of 20,000 iron and "metal wood" clubs a month.

Example Run Through

Activity:

Company:

Product:

Start Date:

Relationship:

Company:

Joint Venture Company:

Activity:

Amount:

Example Run Through

Stage 1: Complex Words

- "set up"
- "trading house"
- "new Taiwan dollars"
- names of people and locations, dates, times, and other basic entities

Example Run Through

Bridgestone Sports Co. said Friday it has set up a **joint venture** in Taiwan with a **local concern** and a Japanese **trading house** to produce **golf clubs** to be shipped to Japan.

The **joint venture**, **Bridgestone Sports Taiwan Co.**, capitalized at 20 million **new Taiwan dollars**, will start production in January 1990 with production of 20,000 iron and "**metal wood**" **clubs** a month.

Example Run Through

Stage 2: Basic Phrases

- Noun group
- Verb group
- Preposition, Conjunction, etc. (Any other syntactic structures that can be identified at this point should go here)

Example Run Through

Stage 2: Basic Phrases (Some Examples)

- approximately 5 kg
- more than 30 people
- the newly elected president
- the largest leftist political force
- a government and commercial project

Example Run Through

Company Name:	Bridgestone Sports Co.
Verb Group	said
Noun Group	Friday
Noun Group	it
Verb Group	had <u>set up</u>
Noun Group	a <u>joint venture</u>
Preposition	in
Location	<u>Taiwan</u>
Preposition	with
Noun Group	a <u>local concern</u>
and	a Japanese <u>trading house</u>
Verb Group	to produce
Noun Group	<u>golf clubs</u>
Verb Group	to be shipped
Preposition	to
Location	<u>Japan</u>

Example Run Through

Stage 3: Complex Phrases

- Attachment of appositives
- Construction of Measure Phrases
- Prepositional attachment of “for” and “of”
- Noun group conjunction

Create relations from complex noun groups

Example Run Through

Stage 3: Complex Phrases (Some Examples)

- The joint venture, Bridgestone Sports Taiwan Co.,
- 20,000 iron and "metal wood" clubs a month
- production of 20,000 iron and "metal wood" clubs a month.
- a local concern and a Japanese trading house

Example Run Through

Bridgestone Sports Co. said Friday it has set up a joint venture in Taiwan with **a local concern and a Japanese trading house** to produce golf clubs to be shipped to Japan.

The joint venture, Bridgestone Sports Taiwan Co., capitalized at 20 million new Taiwan dollars, will start production in January 1990 with **production of 20,000 iron and "metal wood" clubs a month.**

Example Run Through

The joint venture, Bridgestone Sports Taiwan Co.,

Relationship:	TIE-UP
Company:	--
Joint Venture Company:	Bridgestone Sports Taiwan Co.
Activity:	--
Amount:	--

production of 20,000 iron and "metal wood" clubs a month

Activity:	PRODUCTION
Company:	--
Product:	"iron and 'metal wood' clubs
Start Date:	--

Example Run Through

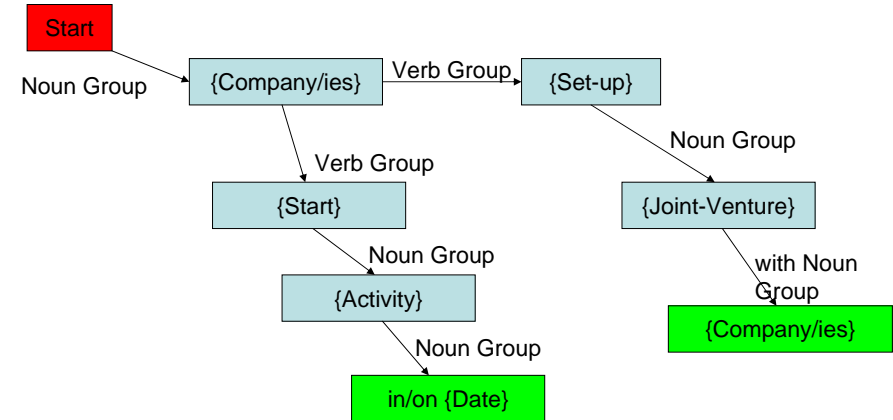
Stage 4: Domain Events

Try to match phrases to the schema

- {Company/ies} {Set-up} {Joint-Venture} with {Company/ies}
- {Produce} {Product}
- {Company} {Capitalized} at {Currency}
- {Company} {Start} {Activity} in/on {Date}

Example Run Through

Stage 4: Domain Events



Example Run Through

Company Name:	Bridgestone Sports Co.	{Company/ies}
Verb Group	said	
Noun Group	Friday	
Noun Group	it	
Verb Group	had <u>set up</u>	{Setup}
Noun Group	a <u>joint venture</u>	{Joint-Venture}
Preposition	in	
Location	Taiwan	
Preposition	with	with
Noun Group	a <u>local concern</u>	{Company/ies}
and	a <u>Japanese trading house</u>	
Verb Group	to produce	{Produce}
Noun Group	<u>golf clubs</u>	{Product}
Verb Group	to be shipped	
Preposition	to	
Location	Japan	

Relationship:	TIE-UP
Entities:	"Bridgestone Sports Co." "a local concern" "a Japanese trading house"
Joint Venture Company:	
Activity:	
Amount:	
Activity:	PRODUCTION
Company:	
Product:	"golf clubs"
Start/Date:	

Example Run Through

Relationship:	TIE-UP
Entities:	"Bridgestone Sports Co." "a local concern" "a Japanese trading house"
Joint Venture Company:	
Activity:	
Amount:	
Activity:	PRODUCTION
Company:	
Product:	"golf clubs"
Start/Date:	

Relationship:	TIE-UP
Entities:	
Joint Venture Company:	"BridgeStone Sports Taiwan Co."
Activity:	
Amount:	NT\$2000000
Activity:	PRODUCTION
Company:	"Bridgestone Sports Taiwan Co."
Product:	
State/Date:	DURING: January 1990

Example Run Through

Stage 5: Merging Structures

So far, 1 or more relation for each sentence.
Relationships spanning multiple sentences?

Coreference?

Example Run Through

Relationship:	TIE-UP	Relationship:	TIE-UP
Entities:	"Bridgestone Sports Co." "a local concern" "a Japanese trading house"	Entities:	
Joint Venture Company:		Joint Venture Company:	"BridgeStone Sports Taiwan Co."
Activity:		Activity:	
Amount:		Amount:	NT\$2000000
Activity:	PRODUCTION	Activity:	PRODUCTION
Company:		Company:	"Bridgestone Sports Taiwan Co."
Product:	"golf clubs"	Product:	
Start Date:		State Date:	DURING: January 1990

Relationship:	TIE-UP	Activity:	PRODUCTION
Company:	--	Company:	--
Joint Venture Company:	Bridgestone Sports Taiwan Co.	Product:	"iron and 'metal wood' clubs"
Activity:	--	Start Date:	--
Amount:	--		

Example Run Through

Activity:	PRODUCTION
Company:	
Product:	"golf clubs"
Start Date:	

Activity:	PRODUCTION
Company:	"Bridgestone Sports Taiwan Co."
Product:	
State Date:	DURING: January 1990

Activity:	PRODUCTION
Company:	--
Product:	"iron and 'metal wood' clubs"
Start Date:	--

Activity:	PRODUCTION
Company:	Bridgestone Sports Co.
Product:	iron and "metal wood" clubs
Start Date:	DURING: January 1990

Example Run Through

Relationship:	TIE-UP	Relationship:	TIE-UP
Entities:	"Bridgestone Sports Co." "a local concern" "a Japanese trading house"	Entities:	Bridgestone Sports Co. a local concern a Japanese trading house
Joint Venture Company:		Joint Venture Company:	Bridgestone Sports Taiwan Co.
Activity:		Activity:	
Amount:		Amount:	NT\$2000000
Relationship:	TIE-UP		
Entities:			
Joint Venture Company:	"BridgeStone Sports Taiwan Co."		
Activity:			
Amount:	NT\$2000000		

Relationship:	TIE-UP
Company:	--
Joint Venture Company:	Bridgestone Sports Taiwan Co.
Activity:	--
Amount:	--

Performance

	1 st blind test	2 nd blind test
Recall	44%	44%
Precision	55%	52%
F-Score	48.9	47.7
GE-Recall	62%	
GE-Precision	53%	
Human	65-80%	

Extensibility?