CSC485 - A4

Michael Dzamba - 994803806

November 23, 2009

NOTE: I am sorry that the formatting is a bit off and that I used probably two more pages then I needed to. I will try to make it up to the planet somehow.

1 Part 1: Finding hypernym relations

1.1 Please see end of assignment for code

1.2 Chunker

The chunker that was used is a modified version of the original chunker. A modification was added to the original NP chunker that allows NP expressions to be of the form, NP (PP NP)*. This helps prevent the chunker from picking up only the last NP in a string of NPs used as prepositional attachments, for use with a predefined hypernym relations. For example, with the original chunker the following would be chunked,

'(NP Dogs) (IN from) (NP Russia) such as (NP Nikola)'

And after using the 'such as' rule from Marti Hearst's paper we arrive at the following relation,

Hyponym(Nikola,Russia)

The modified chunker used would chunk the sentence as,

'(NP Dogs from Russia) such as (NP Nikola)'

Allowing for post processing steps later on to detect the propositional phrase in the first NP and drop the sentence, because of ambiguity of which word(s) the relation is being made for.

1.3 Lexicosyntactic patterns

The implemented Lexicosyntactic patterns are the 6 from Marti Heart's paper and an additional 4 that were mentioned in Jurafsky and Martin from Snow et. al (2005) paper. The last 4 patterns are implemented but commented out, because they produce a large fraction of additional instances, of which many are incorrect.

1.4 Evaluation and report

Confidence level	Instances found	Case 1	Case 2	Case 3	Case 4
Low	336	24	6	86	224
Medium	2	1	0	0	1
High	0	0	0	0	0

There were only two medium confidence relations found, these are,

hyponym(demons, spiritual beings)
This relation is found in case 4, it is there because 'spiritual beings' has no synset in wordnet. Although this relation is perfectly fine according to me.

hyponym(bacteria, microorganisms)
This relation is found in case 1, and is correct.

1.4.1 Suggestions belonging to case 1

Hypernym	Hyponym
attributes	caste
subjects	architecture
metals	iron
condition	dirt
substances	greases
gases	oxygen
organizations	newspapers
substances	particles
substances	oils
jewels	emeralds
object	tree
meats	hamburger
buildings	dormitories

Hypernym	Hyponym
attributes	sex
records	documents
language	russian
animals	mink
jewels	pearls
piping	sewers
buildings	hotels
buildings	houses
records	memoranda
city	atlanta
liquids	water
microorganisms	bacteria

1.4.2 Suggestions belonging to case 2

Hypernym	Hyponym
developments	processes
products	chemicals
developments	uses

Hypernym	Hyponym
submarines	ships
records	documents
fertilizers	materials

1.4.3 Suggestions belonging to case 3

Hypernym	Hyponym	Correct	Comments
diseases	encephalitis	Yes	
industry	irrigation	No	Sentence wrongly chunked
spots	knees	Yes	Very general
ologies	parisology	Yes	
developments	processes	Yes	
nations	peoples	No	
members	hands	Yes	
concepts	blasphemy	Yes	Very general
compensation	salary	Yes	
equipment	dies	No	Dies(dyes) spelt wrong
products	chemicals	Yes	Very general
developments	uses	Yes	
instructions	time	No	
concept	suggestibility	Yes	Very general
concepts	damnation	Yes	Very general
ologies	amphibology	Yes	
subjects	textiles	No	
managers	capitalists	Yes	
provisions	terms	No	
foliage	fern	Yes	
businesses	janitors	No	
submarines	ships	No	Relation backwards
species	dog	Yes	
things	grillwork	Yes	Very general
condition	flies	No	
gases	purposes	No	Badly chunked
members	feet	Yes	
areas	education	Yes	Very general
industry	transport	Yes	
concepts	chain	No	Badly chunked
piping	water	No	
developments	products	Yes	Very general
industry	power	Yes	
girls	sand	No	
reason	economy	No	
nation	switzerland	Yes	

Hypernym	Hyponym	Correct	Comments
subjects	color	No	
passage	following	No	
businesses	factories	Yes	
microorganisms	molds	Yes	
paraphernalia	vehicles	Yes	Very general
people	spouses	Yes	
concepts	purgation	Yes	Very general
corporations	banks	Yes	
pictures	durer	No	
property	facilities	No	
members	head	Yes	
records	correspondence	No	
additives	polyester	Yes	Very general

The above 50 examples from case 3 show 19 examples where the relation is regarded as incorrect and 31 where the relation is regarded as correct. Some of the short comings of my program are the ability to parse more complicated sentence structures. For example *industry* is not a hypernym of *irrigation*, but this is a bit misleading to the original sentence,

(S (NP Industry/NN) ,/, including/IN (NP the/AT production/NN of/IN fertilizer/NN) ,/, (NP irrigation/NN) and/CC (NP power/NN) ,/, (NP transport/NN) and/CC (NP communications/NNS) ,/, and/CC (NP credit/NN institutions/NNS) ;/. ;/.)

I think the sentence is trying to say that *industry* is a hypernym of *the production of irrigation*, which would be a correct statement.
Another short coming is the very general statements made, that although may be considered a hypernym/hyponym relation, they are not all that useful. For example,

The hyponym *education* of hypernym *areas*. Knowing that something is a hyponym of *areas* gives very little information, because the set of hyponyms of *areas* is so large.

This is not a short coming of the program though, but it is a problem with the relation at hand.
Many successes of the program can be seen even though the relation in

wordnet is not present, for example,

The hyponym *encephalitis* of hypernym *diseases*

1.4.4 Suggestions belonging to case 4

Hypernym	Hyponym	Correct	Comments
large sign manufacturers	advance neon sign co	Yes	Very specific
system	fromms	No	Fromm's is a person
societies	american historical association	Yes	Very specific
marine life	clams	Yes	
900calorie milk formulas	quaker oatss quota	Yes	Very specific
necessities	own clothing	Yes	
mighty hand	master	?	Unclear
promotional material	posters	Yes	
taurida	crimean peninsula	No	Was part of Taurida
work	new jazz experiments	No	
pathogenic phenomena	undue regression	Yes	
relevant matters	such sale	No	
intellectuals	aaron copland	Yes	Very specific
radiationproduced diseases	leukemia	Yes	Bad parse
organ meats	chicken liver	Yes	
emotional disturbance	frustration	Yes	
thing	mental illness	Yes	Very general
good cause	illness	Yes	
companies	leesona	Yes	Very specific
heavy material	books	Yes	
hard material	metal	Yes	
organizations	radiotv stations	Yes	Bad parse
promotional material	brochures	Yes	
public facilities	roads	Yes	
policy literature	national security council papers	Yes	Very specific
poor white wine	dry wine	?	Relative to person
small businesses	wholesalers	Yes	
cholesterolrich foods	even eggs	Yes	Bad parse
principle	cousin elec	Yes	Very specific
paraphernalia	architectural developments	Yes	Very general
federal land	national forest	Yes	
small object	cinder	Yes	
professional organizations	various trade associations	Yes	
employment opportunities	regular job	Yes	
diplomats	israeli foreign minister	Yes	
ten largest cities	new york	Yes	Bad parse
such ladies accessories	pistols	No	

Hypernym	Hyponym	Correct	Comments
st petersburg	joe dimaggio	No	Bad parse
25 minutes	san antonio	No	Bad parse
oil mill supplies	belting	Yes	
groups	kiwanis	Yes	
shiny characters	harry truman	Yes	Very specific
region	coastal lowlands	Yes	
organ meats	beef	Yes	
such ladies accessories	automatic weapons	No	
percussive instruments	drums	Yes	
societies	american folklore society	Yes	
civic organizations	local banks	Yes	
sober commentators	nation	No	
reports	publicopinion polls	Yes	Bad parse

The above 50 examples from case 4 show 9 incorrect relations, 2 unclear relations and 39 correct relations. Although a relation may be classified as correct, this does not mean that the relation is useful in a information sense (as described above under case 3 examples). One of the short comings of my program is its ability to parse more complex sentence structures. For example,

(S From/IN (NP International/JJ-TL Airport/NN-TL in/IN Los/NP Angeles/NP to/IN International/JJ-TL Airport/NN-TL in/IN Houston/NP) ,/, as/CS (NP the/AT great/JJ four-jet/JJ Boeing/NP-TL) 707/CD-TL flies/VBZ ,/, is/BEZ (NP a/AT routine/JJ five/CD hours/NNS) and/CC (NP 25/CD minutes/NNS) ,/, including/IN (NP stopovers/NNS at/IN Phoenix/NP) ,/, (NP El/NP Paso/NP) ,/, and/CC (NP San/NP Antonio/NP) ./.)

Is parsed with one of the lexicosyntactic patterns to produce the relation hyponym(san antonio,25 minutes). This is due to the fact that *(NP 25/CD minutes/NNS)* is not chunked together with its larger constituent, and is used as if it were the sentence,

25 minutes, including San antonio

Another short coming can be seen in the sentence,

(S (NP Ralph/NP Houk/NP) ,/, (NP successor/NN to/IN Casey/NP Stengel/NP

at/IN the/AT Yankee/NP helm/NN) ,/, plans/VBZ to/TO bring/VB (NP the/AT entire/JJ New/JJ-TL York/NP-TL squad/NN) here/RB from/IN (NP St./NP Petersburg/NP) ,/, including/IN (NP Joe/NP Dimaggio/NP) and/CC (NP large/JJ crowds/NNS) are/BER anticipated/VBN for/IN both/ABX (NP weekend/NN games/NNS) ./.)

Which gives rise to the relation hyponym(joe dimaggio, st petersburg), when it should really be something like hyponym(joe dimaggio, new york squad member). This is most likely the chunkers fault as it did not chunk *(NP St./NP Petersburg/NP)* with *(NP the/AT entire/JJ New/JJ-TL York/NP-TL squad/NN)* and leave it for additional post processing. Doing this can get tricky though, since propositional phrase attachment is not always straightforward.

Success can be seen from the fact that from the above 50 examples, 39 are regarded by me to be correct (not necessarily useful).

2 Part 2: Find causal relations

2.1 Please see end of assignment for code

np1	v	pp (with/to)	np2	Correct
interpretations	present	None	music	Yes
roar	changed	to	growl	No
orders	ending	None	campaign	Yes
disaster	struck	None	discovery	No
book	went	to	press	No
man	sought	None	justice	Yes
church	meets	None	change	No
owners	met	None	defeat	No
pike	jumped	to	conclusion	No
determination	deserve	None	help	No
children	invited	to	concert	No
bureau	contributed	to	planning	Yes
work	demands	None	attention	Yes
corporation	making	None	acquisition	Yes
food	produces	None	changes	Yes
quest	offers	None	careers	Yes
total	marketing	None	program	No
distress	associated	with	performance	Yes
christianity	contributed	to	education	Yes
grizzlies	winning	None	streak	Yes
screen	drying	None	racks	No
husband	drops	None	case	Yes
company	developed	None	selection	Yes
president	gave	None	word	Yes
fatigues	made	None	streak	No
counterattack	takes	None	line	No
beep	turned	to	buzz	No
palace	issued	None	statement	Yes
president	took	None	pains	Yes
clubs	took	None	line	No
debate	led	to	decision	Yes
teacher	takes	None	hold	Yes
committee	submitted	None	report	Yes
cities	required	None	permits	No
body	held	None	fire	No
playing	gave	None	work	No
age	rewrites	None	events	No

np1	v	pp (with/to)	np2	Correct
issue	lay	with	president	No
group	conducting	None	review	Yes
10000	claiming	None	event	No
gestapo	destroyed	None	rest	Yes
emotions	related	to	fight	Yes
play	running	None	bar	No
president	asks	None	support	Yes
ball	came	to	rest	No
man	made	None	reply	No
will	gives	None	way	Yes
means	make	None	provision	Yes
autumn	dulls	None	blaze	Yes
dilation	accompanying	None	elongation	No

From the above 50 examples, I believe that 26 are correct. In total there were 77 potential causal relations found. From these 68 were formed by the rule,

*(!ABSTRACTION , * , EVENT or ACT)*

One was formed by the rule,

(,lead to,!ENTITY and !GROUP)*

which was, *(debate ,led to, decision)*

also one relation was formed by the rule,

(!ENTITY,associated with,!ABSTRACTION and !GROUP and !POSSESSION)

which was, *(distress, associated with, performance)*

Some short comings of my program were its inability to use more complicated noun phrases as elements in the causal relation. This is because (as mentioned before) it is unclear where propositional phrases attach and which parts of the NP can be dropped in back-off while still keeping the relation in tact.

Also some of the patterns are very general and can include a lot of false-positive relations, this is shown by the large effect of the first rule mentioned with generates 68 out of the 77 potential relations.