Toward a Reading Machine

Anselmo Peñas
UNED NLP & IR Group
nlp.uned.es
Why a Reading Machine?

- The majority of human knowledge is encoded in text
- Much of this text is available in machine-readable formats
- Finding machine operable representations of texts opens the door to the automatic manipulation of vast amounts of knowledge
- There is a big industry awaiting this event.
The goal of Machine Reading

- To build a machine that transforms texts into representations
  - that enable inference and reasoning

- Development of appropriate evaluation methodologies
The goal of Machine Reading

- This goal is not new
  - It is an old dream

- If we failed in the past, why this dream deserves new attention?

- There are two main reasons:
Why now?

1. A change in the paradigms of Computational Semantics
   - Conceptualizing content is not in the form of symbols
   - but in the form of statistical distributions

2. The power to capture the huge amount of background knowledge
   - needed to read a single document
Phase II first attempt

- Questions and Answers are expressed according to a Target Ontology
- The Target Ontology changes with the domain
- Ideally, is an input to the MR system
Query 20011: *Who killed less than 35 people in Kashmir?*

\[\text{:\text{- 'HumanAgent'(FileName, V_y),}\]
\[\text{killingHumanAgent(FileName, V_x, V_y),}\]

\[\text{'HumanAgentKillingAPerson'(FileName, V_x),}\]

\[\text{personGroupKilled(FileName, V_x, V_group),}\]
\[\text{'PersonGroup'(FileName, V_group),}\]
\[\text{'Count'(FileName, V_count),}\]
\[\text{value(FileName, V_count, 35),}\]
\[\text{numberOfMembersUpperBound(FileName, V_group, V_count),}\]

\[\text{eventLocationGPE(FileName, V_x, 'Kashmir').}\]
Representation for reasoning
Machine Reading

- Bridge a gap
- A gap between “representations”

Text

? 

Reasoning System

Answer

Question

Target Ontology

nlp.uned.es
Representation gap

- **Target Ontology** is oriented to express the QA language

- An extension is needed to enable reasoning: **Reasoning Ontology**

- Both are far from text

- We will several intermediate representations
- And a mapping between them
Mapping between representations

Textual Documents

- Reading Representation
- Reasoning Representation
- QA Representation

- Reading Ontology
- Reasoning Ontology
- Target Ontology

- Domain independent
- Domain dependent
RACR NFL v.1

Textual Documents

IE [Roukos et al.]

- Additional Annotated Documents
  - Active Learning?
  - Supervised Learning (re-training for each domain)

Reading Representation

Reasoning Representation

QA Representation

Reading Ontology

Reasoning Ontology

Target Ontology

Entities and relations used by the IE engine

Domain independent

Domain dependent

Inside K-Aggregator [Chalupsky]
Not good enough

- Conjunctive queries
- Performance is product of IE performance for each entity and relation in the query

<table>
<thead>
<tr>
<th>IE F1</th>
<th>Entities and relations in the query</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>'HumanAgent'(FileName, V_y),</td>
</tr>
<tr>
<td>0.8</td>
<td>killingHumanAgent(FileName, V_x, V_y),</td>
</tr>
<tr>
<td>0.9</td>
<td>'HumanAgentKillingAPerson'(FileName, V_x),</td>
</tr>
<tr>
<td>0.8</td>
<td>personGroupKilled(FileName, V_x, V_group),</td>
</tr>
<tr>
<td>0.9</td>
<td>'PersonGroup'(FileName, V_group), 'Count'(FileName, V_count),</td>
</tr>
<tr>
<td></td>
<td>value(FileName, V_count, 35),</td>
</tr>
<tr>
<td></td>
<td>numberOfMembersUpperBound(FileName, V_group, V_count),</td>
</tr>
<tr>
<td>0.9</td>
<td>eventLocationGPE(FileName, V_x, 'Kashmir').</td>
</tr>
<tr>
<td>0.42</td>
<td>Upper bound Performance</td>
</tr>
</tbody>
</table>

- Will the Reasoning Machine recover from that?
A challenge

- Complex queries
- Almost no training data
- Several different domains
- ...

Many big research questions:
  - System architecture, Background Knowledge acquisition, Inference, Representation, ...
Reading Machine
IE approach

- This is not just representation
- Some information is added
  - Learnt during IE training
- Some information is lost
  - Target Ontology
  - Targeted Reading
We need

- Reading representation
- Add information
  - Enrichment: Fill the gaps with Background Knowledge
  - Acquire the Background Knowledge from previous readings
- Targeted reading
  - Closer and closer to the reasoning representation
Bridging the gap

- Reasoning with representations closer to the text
- Representations of text closer to the Reasoning Ontology
- Adding the information needed for reasoning
I can make graphs!

Do you like syntactic dependencies?

I like graphs

Well...

Reasoning Machine

Text

Question

Answer

Reasoning Ontology

Target Ontology
Can we represent Jerry’s Logic Forms as graphs?  
How far would they be from typed syntactic dependencies?  
Can we take advantage of other theories such as Mel'čuk’s Meaning Text Theory?  
Can we have a representation of the whole document instead sentence by sentence?  
Just pictures...
Semantic Dependency Graphs

Let’s call it Semantic Dependency Graph

[Research question #1]
I have entities and classes!

What about person / organization / location / other?

I like entities and classes

Silly machine...

Text → Reading Machine → ? → Reasoning Machine → Answer

Reasoning Ontology

Target Ontology

What else?
Classes

Easy ones

- Entity
  - Named
    - Person
    - Organization
    - Location
    - Other
  - Date
  - Time
  - Measure
    - Distance
    - Weight
    - Height
    - Other
  - ...

Not easy ones

- The rest of words
- (almost)
- Skip all philosophical argue
Can you help me?

Sure! This is about US football

Great...

Reading Machine

Question

Answer

Reasoning Machine

Reasoning Ontology

Target Ontology

Maybe I should **read** something about US football
Classes from text

- Do texts point out classes? Of course
- What classes? The relevant classes for reading
- Uhm... this could be interesting...
- Just small experiment:
  - Parse 30,000 docs. about US football
  - Look for these dependencies
Most frequent has-instance

334: has_instance: [quarterback:n, ('Kerry':'Collins'):name].
306: has_instance: [end:n, ('Michael':'Strahan'):name].
192: has_instance: [team:n, 'Giants':name].
178: has_instance: [owner:n, ('Jerry':'Jones'):name].
151: has_instance: [linebacker:n, ('Jessie':'Armstead'):name].
145: has_instance: [coach:n, ('Bill':'Parcells'):name].
139: has_instance: [receiver:n, ('Amani':'Toomer'):name].
...
### Most frequent classes

<table>
<thead>
<tr>
<th>Class</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>quarterback</td>
<td>15457</td>
</tr>
<tr>
<td>coach</td>
<td>12395</td>
</tr>
<tr>
<td>end</td>
<td>7865</td>
</tr>
<tr>
<td>receiver</td>
<td>7611</td>
</tr>
<tr>
<td>linebacker</td>
<td>6794</td>
</tr>
<tr>
<td>receiver</td>
<td>3479</td>
</tr>
<tr>
<td>(defensive:receiver)</td>
<td></td>
</tr>
<tr>
<td>(wide:receiver)</td>
<td></td>
</tr>
<tr>
<td>team</td>
<td>3265</td>
</tr>
<tr>
<td>coordinator</td>
<td>3252</td>
</tr>
<tr>
<td>(defensive:coordinator)</td>
<td></td>
</tr>
<tr>
<td>(offensive:coordinator)</td>
<td></td>
</tr>
<tr>
<td>player</td>
<td>2870</td>
</tr>
<tr>
<td>agent</td>
<td>2790</td>
</tr>
<tr>
<td>guard</td>
<td>2291</td>
</tr>
<tr>
<td>pick</td>
<td>2258</td>
</tr>
<tr>
<td>manager</td>
<td>2177</td>
</tr>
<tr>
<td>(head:coach)</td>
<td></td>
</tr>
<tr>
<td>rookie</td>
<td>2082</td>
</tr>
<tr>
<td>back</td>
<td>2039</td>
</tr>
<tr>
<td>manager</td>
<td>1985</td>
</tr>
<tr>
<td>(defensive:coordinator)</td>
<td></td>
</tr>
<tr>
<td>(offensive:coordinator)</td>
<td></td>
</tr>
<tr>
<td>lineman</td>
<td>1885</td>
</tr>
<tr>
<td>(general:manager)</td>
<td>1776</td>
</tr>
<tr>
<td>(vice:president)</td>
<td>1425</td>
</tr>
<tr>
<td>(offensive:coordinator)</td>
<td></td>
</tr>
<tr>
<td>(free:agent)</td>
<td>1366</td>
</tr>
<tr>
<td>champion</td>
<td>1196</td>
</tr>
<tr>
<td>(head:coach)</td>
<td></td>
</tr>
<tr>
<td>(vice:president)</td>
<td></td>
</tr>
<tr>
<td>(free:agent)</td>
<td></td>
</tr>
<tr>
<td>spokesman</td>
<td>1140</td>
</tr>
<tr>
<td>(head:coach)</td>
<td></td>
</tr>
<tr>
<td>(vice:president)</td>
<td></td>
</tr>
<tr>
<td>(free:agent)</td>
<td></td>
</tr>
<tr>
<td>champion</td>
<td>1102</td>
</tr>
<tr>
<td>(head:coach)</td>
<td></td>
</tr>
<tr>
<td>(vice:president)</td>
<td></td>
</tr>
<tr>
<td>(free:agent)</td>
<td></td>
</tr>
<tr>
<td>manager</td>
<td>1006</td>
</tr>
<tr>
<td>(head:coach)</td>
<td></td>
</tr>
<tr>
<td>(vice:president)</td>
<td></td>
</tr>
<tr>
<td>(free:agent)</td>
<td></td>
</tr>
<tr>
<td>champion</td>
<td>990</td>
</tr>
<tr>
<td>(head:coach)</td>
<td></td>
</tr>
<tr>
<td>(vice:president)</td>
<td></td>
</tr>
<tr>
<td>(free:agent)</td>
<td></td>
</tr>
<tr>
<td>manager</td>
<td>989</td>
</tr>
<tr>
<td>(head:coach)</td>
<td></td>
</tr>
<tr>
<td>(vice:president)</td>
<td></td>
</tr>
<tr>
<td>(free:agent)</td>
<td></td>
</tr>
<tr>
<td>champion</td>
<td>987</td>
</tr>
<tr>
<td>(head:coach)</td>
<td></td>
</tr>
<tr>
<td>(vice:president)</td>
<td></td>
</tr>
<tr>
<td>(free:agent)</td>
<td></td>
</tr>
<tr>
<td>tailback</td>
<td>987</td>
</tr>
</tbody>
</table>
Find more ways to point out a class in text?

- Now you have thousands of seeds
- Bootstrap!

[Research question #2]
I have these classes!

What?

What a mess! Flat, redundant...! Where are gameWinner, safetyPartialCount, ...?

Text → Reading Machine → Reasoning Machine

Question → Answer

Reasoning Ontology

Target Ontology
Ok, let’s move on... Show me the relations you have.

gameWinner a class!? Come on!
Tom, ehm... What’s a relation?

Well... certainly, a relation is a n-tuple...

... n-tuple...

Like verb structures?

Uhm... this could be interesting...

Just small experiment:

- Take the 30,000 docs. about US football and look for:
Most frequent propositions

nvn: [person:n, do:v, thing:n]: 1530.
nvn: [group:n, do:v, thing:n]: 1264.
nvn: [group:n, win:v, game:n]: 960.
nvn: [person:n, tell:v, person:n]: 902.
nvn: [person:n, catch:v, pass:n]: 814.
nvn: [group:n, have:v, chance:n]: 656.
nvn: [person:n, miss:v, game:n]: 580.
nvn: [person:n, throw:v, pass:n]: 567.
nvn: [group:n, have:v, lot:n]: 551.
nvn: [person:n, do:v, job:n]: 490.
nvn: [group:n, lose:v, game:n]: 482.
nvn: [person:n, have:v, problem:n]: 479.
nvn: [person:n, tell:v, group:n]: 473.
nvn: [person:n, throw:v, interception:n]: 465.
nvn: [group:n, play:v, game:n]: 464.

Considering pronouns
Most frequent propositions

nvn:[team:n, win:v, game:n]:297.

nvn:[team:n, have:v, record:n]:212.

nvn:[team:n, lose:v, game:n]:166.

nvn:[touchdown:n, give:v, lead:n]:160.

nvn:[team:n, play:v, game:n]:154.

nvn:[goal:n, give:v, lead:n]:154.

nvn:[(field:goal):n, give:v, lead:n]:150.

nvn:[team:n, make:v, playoff:n]:146.

nvn:[team:n, win:v, championship:n]:136.

nvn:[touchdown:n, give:v, ((0: - : 0):lead):n]:135.

nvn:[goal:n, give:v, ((0: - : 0):lead):n]:124.

nvn:[(field:goal):n, give:v, ((0: - : 0):lead):n]:123.

nvn:[offense:n, score:v, touchdown:n]:118.
Most frequent “relations”

- A **relation** is a n-tuple between instances of certain type
  - I was playing with NNPs as instances...

- Let’s generalize and aggregate:
  - Marino -> NAME
  - Bulger -> NAME
  - Jones -> NAME
  - ...
Most frequent “relations”

- nvn: ['NAME', play:v, 'NAME']: 2382.
- nvn: ['NAME', lead:v, 'NAME']: 2285.
- nvn: ['NAME', score:v, touchdown:n]: 2261.
- nvn: ['NAME', have:v, 'NAME']: 1866.
- nvn: ['NAME', kick:v, goal:n]: 1433.

Now
- I know instance classes
- I know propositions between instances
- Let’s find the probability of the classes given a proposition
Most probable typed relations

nvn:['NAME', throw:v, pass:n]:[quarterback]:0.00116408907425231.
nvn:['NAME', catch:v, pass:n]:[receiver]:0.000947483419496505.
nvn:['NAME', throw:v, (touchdown:pass):n]:[quarterback]:0.000767205964030132.
nvn:['NAME', throw:v, interception:n]:[quarterback]:0.000415367155661766.
nvn:['NAME', catch:v, pass:n]:[(wide:receiver)]:0.000406047565796885.
nvn:['NAME', say:v, linebacker:n]:[linebacker]:0.000395901727835594.
nvn:['NAME', rank:v, (no:'.'):n]:[(no:'.')]:0.000392502869262291.
nvn:['NAME', rank:v, (0:no:'.'):n]:[(no:'.')]:0.000392502869262291.
nvn:['NAME', complete:v, pass:n]:[quarterback]:0.000390269288924688.
nvn:['NAME', catch:v, (0:pass):n]:[receiver]:0.000346249735543358.
nvn:['NAME', catch:v, pass:n]:[end]:0.0003339730214679.
nvn:['NAME', throw:v, ((0: - : yard):pass):n]:[quarterback]:0.000304392988179183.
nvn:['NAME', have:v, sack:n]:[end]:0.000297055408474171.
nvn:['NAME', intercept:v, pass:n]:[safety]:0.000292011905223431.
Relations between classes

Now I can ask about the relations between classes

Quarterback & receiver

nvn:['NAME', hit:v, 'NAME'][quarterback, receiver]:3.67432997918756e-06.
nvn:['NAME', find:v, 'NAME'][quarterback, receiver]:1.8192935712796e-06.
nvnpn:['NAME', complete:v, pass:n, to:in, 'NAME'][quarterback, receiver]:1.4512783860507e-06.
nvnpn:['NAME', throw:v, pass:n, to:in, 'NAME'][quarterback, receiver]:1.37642726590848e-06.
nvnpn:['NAME', catch:v, pass:n, from:in, 'NAME'][receiver, quarterback]:1.16492444555009e-06.
nvnpn:['NAME', throw:v, (touchdown:pass):n, to:in, 'NAME'][quarterback, receiver]:1.0606850217847e-06.

If there is only one relevant relation between them, they are paraphrases
But you can have many relevant relations between same classes

Team and game

nvn:['NAME', win:v, game:n]:[team]:9.69097313067351e-05.
nvn:['NAME', lose:v, game:n]:[team]:5.96928038789563e-05.
nvn:['NAME', play:v, game:n]:[team]:2.7232092783388e-05.
nvn:['NAME', have:v, game:n]:[team]:2.5404025345459e-05.
nvn:['NAME', enter:v, game:n]:[team]:1.32010748425686e-05.
nvn:['NAME', (not:win)v, game:n]:[team]:5.77931930517973e-06.
nvn:['NAME', forfeit:v, game:n]:[team]:5.30734251201793e-06.
nvn:['NAME', tie:v, game:n]:[team]:3.95409472849798e-06.
nvn:['NAME', reach:v, game:n]:[team]:3.66152627590672e-06.
nvn:['NAME', average:v, game:n]:[team]:3.45676070239657e-06.
nvnpn:['NAME', extend:v, streak:n, to:in, game:n]:[team]:3.29470336174047e-06.

How to cluster different realizations of the same relation?

[Research question #3]
Now, when I see something (named) doing things, I can guess its class

Culpepper directed a 15-play drive

- nvn:['NAME', direct:v, drive:n]:[quarterback]:3.37813951038364e-05.
- nvn:['NAME', direct:v, drive:n]:[backup]:2.98954603541518e-06.
- nvn:['NAME', direct:v, drive:n]:[man]:1.56144396948542e-06.
- nvn:['NAME', direct:v, drive:n]:[freshman]:1.48171220798502e-06.
- nvn:['NAME', direct:v, drive:n]:[passer]:1.3913157132247e-06.

And when I see it doing many things in a document, I can aggregate the evidence on its class. How?

[Research question #4]
Propositions into axioms

- “Can we axiomatize this?”, asked Jerry
- Why not?
- \( P(\text{quarterback, throw, pass}) = 0.0011 \)
  - \( P(\text{quarterback} \mid \text{throw, pass}) = p \)
  - \( \text{throw}(x,y), \text{pass}(y) \rightarrow \text{quarterback}(x) \mid p \)
Relations between events

- “Can you find the ways to express causality?”, asked Rutu

- Why not? Give me a seed

Touchdown & victory

NVN 14 'touchdown':'give':'victory'
NVN 11 'touchdown':'seal':'victory'
NVN 4 'touchdown':'secure':'victory'
 Relations between events

Give

NVN 136 'touchdown':'give':'lead'
NVN 130 'goal':'give':'lead'
NVN 85 'pass':'give':'lead'

Seal

NVN 12 'interception':'seal':'victory'
NVN 11 'touchdown':'seal':'victory'
NVN 6 'pass':'seal':'victory'

Secure

NVN 5 'victory':'secure':'title'
NVN 4 'group':'secure':'title'
NVN 4 'victory':'secure':'championship'
Relations between events

Set_up

NVN 25 'interception':'set_up':'touchdown'
NVN 20 'pass':'set_up':'touchdown'
NVN 19 'interception':'set_up':'goal'
NVN 14 'pass':'set_up':'goal'
NVN 14 'return':'set_up':'touchdown'
NVN 12 'interception':'set_up':'score'
NVN 11 'fumble':'set_up':'touchdown'
NVN 11 'run':'set_up':'touchdown'
NVN 10 'person':'set_up':'touchdown'
NVN 9 'return':'set_up':'goal'
NVN 9 'pass':'set_up':'run'
NVN 9 'interception':'set_up':'run'
NVN 9 'run':'set_up':'goal'
IE approach

Textual Documents → IE → Reading Representation → Entities and relations used by the IE engine

Reading Machine?
The Reading Machine

Textual Documents → Reading → Aggregation and generalization

Rumination Cycle

Enrichment → Mapping → Background Knowledge

Seeding

Target Ontology

Reasoning Machine

Reasoning Ontology

Question → Answer
The Reading Machine

Iterate over big amounts of text

1. Process texts up to reified logic forms
2. Generalization and aggregation of frequent structures in the logic forms
   1. Estimation of their probability distributions
   2. Dumping intermediate Background Knowledge Bases (BKB)
3. Enrichment of initial text representations (BKBs are then used to fill the knowledge gaps)
4. Repeat with the enriched representation of texts
The Reading Machine

Textual Documents → Reading

Question → Reading Machine → Answer

Reasoning Machine

Target Ontology

Reasoning Ontology

Aggregation and generalization

Rumination Cycle

Enrichment

Mapping

Background Knowledge

Seeding
...to set up a 7-yard Young touchdown pass to Brent Jones

Young pass

?

> X:has-instance:Young
  X=quarterback

> NVN:quarterback:X:pass
  X=throw
  X=complete

pass to Jones

?

> X:has-instance:Jones
  X=end

> NVN:end:X:pass
  X=catch
  X=drop
... to set up a 7-yard Young touchdown pass to Brent Jones

touchdown pass

?> NVN  touchdown:X:pass
  False
?> NPN  pass:X:touchdown
  X=for
...to set up a 7-yard Young touchdown pass to Brent Jones

```plaintext
?> NVNPN  NAME:X:pass:for:touchdown
  X=complete
  X=catch
```
...to set up a 7-yard Young touchdown pass to Brent Jones

⇒ Young complete pass for touchdown
⇒ Jones catch pass for touchdown
Evaluation

- The evaluation of these machines is an open research question.

- It is also an opportunity to consolidate the community around the field.
  - Specially in EU.

- The development of an international evaluation is part of this proposal.
Conclusions

- Machine Reading is a target for CL
- A dream
- Time for a new attempt
  - Concepts are probability distributions
    - Propositions
- New project at UNED
  - Development of a Reading Machine
  - Collect sources of knowledge (text + ...)
  - New international evaluation campaign
References


Thanks!
Domain issue

?> person:X:pass

NFL Domain

905:nvn:[person:n, catch:v, pass:n].
667:nvn:[person:n, throw:v, pass:n].
286:nvn:[person:n, complete:v, pass:n].
204:nvnpn:[person:n, catch:v, pass:n, for:in, yard:n].
85:nvnpn:[person:n, catch:v, pass:n, for:in, touchdown:n].

IC Domain

6:nvn:[person:n, have:v, pass:n]
3:nvn:[person:n, see:v, pass:n]
1:nvnpn:[person:n, wear:v, pass:n, around:in, neck:n]

BIO Domain

<No results>
Domain issue

?> X:receive:Y

NFL Domain
55:nvn:[person:n, receive:v, call:n].
34:nvn:[person:n, receive:v, offer:n].
33:nvn:[person:n, receive:v, bonus:n].
29:nvn:[team:class, receive:v, pick:n].

IC Domain
78 nvn:[person:n, receive:v, call:n]
44 nvn:[person:n, receive:v, letter:n]
35 nvn:[group:n, receive:v, information:n]
31 nvn:[person:n, receive:v, training:n]

BIO Domain
24 nvn:[patients:n, receive:v, treatment:n]
14 nvn:[patients:n, receive:v, therapy:n]
13 nvn:[patients:n, receive:v, care:n]
Limiting to a specific domain provides some powerful benefits:
- Ambiguity is reduced
- Higher density of relevant propositions
- Different distribution of propositions across domains
- Amount of source text is reduced, allowing deeper processing such as parsing
- Specific tools for specific domains

Proposition stores seem to be useful:
- Improve parsing, corref, WSD,...

We presented a new application: ENRICHMENT
Cornerback & receiver

\[
\begin{align*}
nvn(('NNP':'receiver'):'beat':('NNP':'cornerback')) & : 23. \\
nvn(('NNP':'cornerback'):'cover':('NNP':'receiver')) & : 11. \\
nvn(('NNP':'cornerback'):'shadow':('NNP':'receiver')) & : 4. \\
nvn(('NNP':'cornerback'):'shut\_down':('NNP':'receiver')) & : 4. \\
nvn(('NNP':'receiver'):'burn':('NNP':'cornerback')) & : 3. \\
nvn(('NNP':'cornerback'):'intercept':('NNP':'receiver')) & : 3. \\
nvn(('NNP':'cornerback'):'face':('NNP':'receiver')) & : 2. \\
nvn(('NNP':'cornerback'):'trip\_up':('NNP':'receiver')) & : 2. \\
nvn(('NNP':'cornerback'):'hold':('NNP':'receiver')) & : 2. \\
nvn(('NNP':'cornerback'):'outjump':('NNP':'receiver')) & : 2. \\
nvn(('NNP':'cornerback'):'limit':('NNP':'receiver')) & : 2.
\end{align*}
\]
A quarterback is not a player

Players

- nvn('NNP':'player':'catch':'pass'): 83.
- nvn('NNP':'player':'miss':'game'): 66.
- nvn('NNP':'player':'have':'yard'): 59.
- nvn('NNP':'player':'gain':'yard'): 49.
- nvn('NNP':'player':'throw':'pass'): 43.
- nvn('NNP':'player':'score':'touchdown'): 43.
- nvn('NNP':'player':'have':'sack'): 34.
- nvn('NNP':'player':'average':'yard'): 30.
- nvn('NNP':'player':'have':'surgery'): 28.
- nvn('NNP':'player':'run':'yard'): 27.
- nvn('NNP':'player':'start':'game'): 26.
- nvn('NNP':'player':'carry':'ball'): 26.
They are what they do

Quarterbacks

- \text{nvn}((\text{'NNP'}:'quarterback'):'throw':'pass'):1093.
- \text{nvn}((\text{'NNP'}:'quarterback'):'complete':'pass'):422.
- \text{nvn}((\text{'NNP'}:'quarterback'):'throw':'interception'):407.
- \text{nvn}((\text{'NNP'}:'quarterback'):'lead':((\text{'NNP'}:'team'))):354.
- \text{nvn}((\text{'NNP'}:'quarterback'):'start':'game'):211.
- \text{nvn}((\text{'NNP'}:'quarterback'):'replace':((\text{'NNP'}:'quarterback'))):208.
- \text{nvn}((\text{'NNP'}:'quarterback'):'hit':((\text{'NNP'}:'receiver'))):177.
- \text{nvn}((\text{'NNP'}:'quarterback'):'have':'pass'):142.
Rephrasing

“Culpepper finds Moss” <-> “Culpepper completes pass to Moss”

1. Hypothesize Entity classes

?> NAME find NAME

nvn:['NAME', find:v, 'NAME']:[quarterback, receiver]:1.8192935712796e-06
nvn:['NAME', find:v, 'NAME']:[quarterback, end]:8.17538936560293e-07
nvn:['NAME', find:v, 'NAME']:[quarterback, (wide:receiver)]:7.94094409104393e-07
nvn:['NAME', find:v, 'NAME']:[quarterback, (tight:end)]:7.41880417507334e-07

(in Logic Form)

?> find(e,x1,x2), name(x1,NAME), name(x2,NAME) -> class(x1,c1), class(x2,c2)

c1=quarterback, c2=receiver | Prob(e,c1,c2)=1.8192935712796e-06

Prob(e,c1,c2)=8.17538936560293e-07

c1=quarterback, c2=end

c1=quarterback, c2=(wide:receiver) | Prob(e,c1,c2)=7.94094409104393e-07

c1=quarterback, c2=(tight:end) | Prob(e,c1,c2)=7.41880417507334e-07
2. Recover other propositions between the classes found

Propositions relating names of *quarterbacks* with names of *receivers*

\[
\text{nnpn}:['\text{Name}', \text{to}:\text{in}, '\text{Name}']:[\text{quarterback}, \text{receiver}]:5.69990315934166e-06 \\
\text{nvn}:['\text{Name}', \text{hit}:v, '\text{Name}']:[\text{quarterback}, \text{receiver}]:3.67432997918756e-06 \\
\text{nvn}:['\text{Name}', \text{find}:v, '\text{Name}']:[\text{quarterback}, \text{receiver}]:1.8192935712796e-06 \\
\text{nvnpn}:['\text{Name}', \text{throw}:v, \text{to}:\text{in}, '\text{Name}']:[\text{quarterback}, \text{receiver}]:1.7055936330989e-06 \\
\text{nvnpn}:['\text{Name}', \text{complete}:v, \text{pass}:n, \text{to}:\text{in}, '\text{Name}']:[\text{quarterback}, \text{receiver}]:1.4512783860507e-06 \\
\text{nvnpn}:['\text{Name}', \text{throw}:v, \text{pass}:n, \text{to}:\text{in}, '\text{Name}']:[\text{quarterback}, \text{receiver}]:1.37642726590848e-06 \\
\text{nvnpn}:['\text{Name}', \text{catch}:v, \text{pass}:n, \text{from}:\text{in}, '\text{Name}']:[\text{receiver}, \text{quarterback}]:1.16492444555009e-06 \\
\text{nvnpn}:['\text{Name}', \text{throw}:v, (\text{touchdown}:\text{pass}):n, \text{to}:\text{in}, '\text{Name}']:[\text{quarterback}, \text{receiver}]:1.0606850e-06 \\
\text{nvnpn}:['\text{Name}', \text{pass}:v, \text{to}:\text{in}, '\text{Name}']:[\text{quarterback}, \text{receiver}]:9.5383721790507e-07 \\
\text{nvnpn}:['\text{Name}', \text{connect}:v, \text{with}:\text{in}, '\text{Name}']:[\text{quarterback}, \text{receiver}]:8.9380301215876e-07 \\
\text{nnpn}:['\text{Name}', \text{on}:\text{in}, '\text{Name}']:[\text{receiver}, \text{quarterback}]:7.40714092598897e-07 \\
\text{nvnpn}:['\text{Name}', \text{look}:v, \text{for}:\text{in}, '\text{Name}']:[\text{quarterback}, \text{receiver}]:7.16122012330255e-07
\]