Are community-built ontologies robust enough to build maintainable NLG systems?

Pablo Duboue

Textualization Software Ltd.
Vancouver, Canada

Facultad de Matemática, Astronomía y Física
Universidad Nacional de Córdoba
Córdoba, Argentina

NLG Needs Ontological Resources

- Traditionally, NLG needs ontological resources
- Large scale $\Rightarrow$ expensive
- Community-built $\Rightarrow$ unreliable?

Example: Simple NLG Task

- Build the full name of a person
  - Only given names and last name?
- However:
  - Some cultures use paternal last name then maternal last name
  - Others use maternal last name then paternal
  - Others do gender agreement with the paternal last name
- Data needs:
  - Given names
  - Paternal last name
  - Maternal last name
  - Region of origin
  - Gender

Wikipedia to the Rescue

- Some of that info can be scrapped from Wikipedia
  - Such scrapping effort is facilitated by the Infoboxes, the boxes at the top of a page
- A large scale effort to achieve this is DBpedia
- But each triple extracted involves 4 people working somewhat independently:
  - Page editor/author
  - Infobox editor/author
  - DBpedia mapping editor/author
  - DBpedia extractor programmer

This talk

- Community driven resources change dangerously
- Ontological resources might be at the cross-roads
  - Compared to, for example, lexical resources
- Do we need robust ontology-driven systems?
  - Yes.
- Some DBpedia anecdotes and three papers

About the Speaker

- Columbia University – NLG
  - PhD Thesis: “Indirect Supervised Learning of Strategic Generation Logic”, defended Jan. 2005.
- IBM Research Watson – Question Answering
  - Deep QA - Watson: ML component used in the show
- Montreal (Canada) – Consulting
  - Collaboration with Université de Montreal: 1-Click Search
  - Free Software and consulting for Startups and SMBs
- White Plains (NY) – Research sabbatical
  - Personal sabbatical focusing in Research & Free Software
  - Launching a NLG venture in Vancouver
- Robust NLG through human-understandable ML

DBpedia

- DBpedia [Bizer et al., 2009] is an ontology curated from Wikipedia infoboxes
  - Infoboxes are the small tables containing structured information at the top of most Wikipedia pages.
  - The mappings between the infoboxes labels to the ontology is done in a wiki itself: `http://mappings.dbpedia.org/`.
  - The source code of the scrapping scripts is also available with all its development history.
- Not to be confused with a new project targeting to provide structured information to Wikipedia, wikidata.

Infobox ⟺ DBpediaMappings

```
{{About|the 43rd President of the United States|his father, the 41st Presic
H. W. Bush|the American settler|George Washington Bush}}
<!--See [[WP:EDN]]-->
{{Pp-move-indef}}
{{Active editnotice}}{{Pp-semi-blp|small=yes}}
{{Use mdy dates|date=November 2016}}
{{Infobox president
|name          = George W. Bush
|office        = [[List of Presidents of the United States|43rd President c
United States]]
|image         = George-W-Bush.jpeg
|predecessor   = [[Bill Clinton]]
|successor     = [[Barack Obama]]
|vicepresident = [[Dick Cheney]]
|order2        = [[List of Governors of Texas|46th Governor of Texas]]
|lieutenant2   = {{Ublist|[[Bob Bullock]]|Rick Perry}}
|predecessor2  = [[Ann Richards]]
|successor2    = [[Rick Perry]]
|birth_name    = George Walker Bush
|birth_date    = {{Birth date and age|1946|7|6}}
|birth_place   = {{nowrap|[[New Haven, Connecticut]], U.S.}}
|party         = [[Republican Party (United States)|Republican]]
|spouse        = {{Marriage|[[Laura Bush|Laura Welch]]|November 5, 1977}}
|relations     = ''See [[Bush family]]''
|children      = {{Hlist|[[Barbara Bush (born 1981)|Barbara]]|[[Jenna Bush
```

dbpedia.org/page/George_W._Bush

DBpedia · Browse using · Formats

crisis.Bush left office in 2009, and h
a home in a suburban area of Dalla
written a memoir, Decision Points. H
(en)

dbo:almaMater	dbr:Yale_College dbr:Harvard_Business_School
dbo:award	dbr:United_States_Aviator_Badge dbr:National_Defense_Service_Med dbr:Air_Force_Outstanding_Unit_Aw dbr:Marksmanship_Ribbon
dbo:birthDate	1946-07-06 (xsd:date)
dbo:birthPlace	dbr:New_Haven,_Connecticut
dbo:child	dbr:Jenna_Bush_Hager dbr:Barbara_Pierce_Bush
dbo:country	dbr:United_States
dbo:lieutenant	dbr:Rick_Perry dbr:Bob_Bullock

Infobox ⟺ Mappings

```
{{About|the 43rd President of the United States|his father, the 41st Presic
H. W. Bush|the American settler|George Washington Bush}}
<!--See [[WP:EDN]]-->
{{Pp-move-indef}}
{{Active editnotice}}{{Pp-semi-blp|small=yes}}
{{Use mdy dates|date=November 2016}}
{{Infobox president
|name          = George W. Bush
|office        = [[List of Presidents of the United States|43rd President c
United States]]
|image         = George-W-Bush.jpeg
|predecessor   = [[Bill Clinton]]
|successor     = [[Barack Obama]]
|vicepresident = [[Dick Cheney]]
|order2        = [[List of Governors of Texas|46th Governor of Texas]]
|lieutenant2   = {{Ublist|[[Bob Bullock]]|Rick Perry}}
|predecessor2  = [[Ann Richards]]
|successor2    = [[Rick Perry]]
|birth_name    = George Walker Bush
|birth_date    = {{Birth date and age|1946|7|6}}
|birth_place   = {{nowrap|[[New Haven, Connecticut]], U.S.}}
|party         = [[Republican Party (United States)|Republican]]
|spouse        = {{Marriage|[[Laura Bush|Laura Welch]]|November 5, 1977}}
|relations     = ''See [[Bush family]]''
|children      = {{Hlist|[[Barbara Bush (born 1981)|Barbara]]|[[Jenna Bush
```

Mapping en:Infobox president

This is the mapping for the Wikipedia template Infobox president. Find usages of this Wikipedia template here.

Test this mapping (or in namespace File or Creator) with some example Wikipedia pages. Check which properties are not mapped yet.

Read more about mapping Wikipedia templates.

Template Mapping (help)	
map to class	President

Mappings

Property Mapping (help)	
template property	otherparty
ontology property	otherParty

Property Mapping (help)	
template property	name
ontology property	foaf:name

Property Mapping (help)	
template property	birth_date
ontology property	birthDate

Two Versions: Compared

Type files analysis

Property	3.6	2014
Number of triples	6,173,940	28,031,852
Unique subjects (entities)	1,668,503	4,218,628
Unique objects (types)	250	547
Max objects per subject	6	16

Mapping files analysis

Property	3.6	2014
Number of verbs	1,100	1,370
Number of triples	13,795,664	33,449,633

- However, many entities lost their types
  - From 20,693 Politicians in 3.6, 4,542 are gone (20%-25%).
  - However, the total Politicians in 2014 is 40,343.

DBpedia Details

- Changelogs: `http://oldwiki.dbpedia.org/Changelog`
  - DBpedia 2014 (09/2014)
  - DBpedia 3.9 (09/2013)
  - DBpedia 3.8 (08/2012)
  - DBpedia 3.7 (08/2011)
  - DBpedia 3.6 (01/2011)
- DBpedia 3.9 (09/2013): Changelog
  - Core Framework: refined rules for URIs of sub-resources, e.g., for Wikipedia pages having multiple infoboxes
  - These are the type of disruptive changes that you can expect
- But there's more!
  - Current infobox for GWB is president
  - 2014 was officeholder: `https://en.wikipedia.org/w/index.php?title=George_W._Bush&action=edit&oldid`
  - 2011 was president: `https://en.wikipedia.org/w/index.php?title=George_W._Bush&action=edit&oldid=4`

What Might Mean to NLG Practitioners

- Either take a version and freeze it in time
  - Losing the ability to cope with large amount of up-to-date entities
- Be ready to adapt resources for each new version
  - Extra resources built on top might become stale
- For example:
  - A system that has a generation lexicon for Politician will need to be updated for OfficeHolder
  - Some entities will be Politican, some will be OfficeHolder
  - This level of resilience is unusual for NLG

Three Papers

- NAACL 2012
  - DBpedia is useful enough for referring expressions
- MICAI2015/WebNLG2016
  - Using two versions of DBpedia allows studying impact of errors in referring expression genration
- Iberamia 2016
  - Robustness helps to learn Preference Ordering for properties for the Incremental Algorithm
  - Conference is next week!

Collaborators

Martin Dominguez

Paula Estrella

Fabian Pacheco

Referring Expression Generation (REG)

- Classic NLG problem
  - **Input:** set of entities (with a distinguished element), set of triples pertaining to the entities.
  - **Output:** a Definite Description, i.e., a set of *positive triples* and *negative triples*.
  - Focus on running time **efficiency** and generating **succint** and **easily understandable** expressions.

- Example output
  - Task: {‘Eben_Moglen’(EB), ‘Lawrence_Lessig’(LL), ‘Linus_Torvalds’(LT)}

Referent	Incremental Algorithm	Gardent
EB	{ (EB *occupation* Software_Freedom_Law_Center) }	{ (EB *occupation* Software_Freedom_Law_Center) }
LL	{ (LL *birthPlace* United_States), (LL, *occupation* Harvard_Law_School) }	{ (LL *birthPlace* Rapid_City,_South_Dakota) }
LT	{ (LT *occupation* Software_engineer) }	{ (LT *nationality* Finnish_American) }

Incremental Algorithm (IA) – an established REG algo

- Introduced in [Dale and Reiter, 1995]
  - Greedy approach, use a **default ordering**: Preference Order (PO)
  - Iterates over PO and selects a type
  - Adds a triple of the given type one at a time
  - Removes from the confusor set C all entities ruled out by the new triple
  - Triples that do not eliminate any new entity from C are ignored
  - The algorithm terminates when C is empty.
- Many other algorithms
  - Graph
  - Full Brevity
  - Gardent's

Possible Application To Multi-document Summarization

Use REG to fix anaphoric references drafted from different documents (similar to [Siddharthan et al., 2011])

- Excerpt from Columbia Newsblaster:

*Thousands of cheering, flag-waving Palestinians gave Palestinian Authority President Mahmoud Abbas an enthusiastic welcome in Ramallah on Sunday, as he told them triumphantly that a "Palestinian spring" had been born following his speech to the United Nations last week. The **president** pressed Israel, in unusually frank terms, to reach a final peace agreement with the Palestinians, citing the boundaries in place on the eve of the June 1967 Arab-Israeli War as the starting point for negotiation about borders.*

Can REG Help Summarization?

Pacheco, Duboue, Dominguez. *On the feasibility of open domain referring expression generation using large scale folksonomies.* NAACL 2012.

- Do we have data for the relevant entities?
  - Yes, roughly 50% of the time.
  - We used anaphora training data and looked it up on DBpedia by hand.
- Do we have **discriminant** data for relevant entities?
  - Yes, roughly 80% of the time.
  - Measured on Wikinews, Cohen's κ of 79%.
- Are classic REG algorithms enough?
  - *Maybe not,* they either fail to produce an output or return a poor description in 60%+ of the cases.

Experiments With Wikinews-derived REG Tasks

- Wikinews, a news service operated as a wiki
  - Entities disambiguated by *interwiki* links.

```
Former [[New Mexico]] {{w|Governor of New
Mexico|governor}} {{w|Gary Johnson}} ended his
campaign for the {{w|Republican Party (United
States)|Republican Party}}
```

- Human-written Property Ordering:

TYPE ORDERINOFFICE NATIONALITY COUNTRY PROFESSION BIRTHPLACE LEADERNAME^{-1} KEYPERSON^{-1} AUTHOR^{-1} COMMANDER^{-1} OCCUPATION KNOWNFOR INSTRUMENT SUCCESSOR MONARCH SUCCESSOR^{-1} PRIMEMINISTER^{-1} ACTIVEYEARSENDDATE PARTY DEATHDATE DEATHPLACE CHILD ALMAMATER ACTIVEYEARSSTARTDATE RELIGION SPOUSE PRESIDENT^{-1} NOTABLECOMMANDER^{-1} VICEPRESIDENT PRESIDENT PRIMEMINISTER AWARD MILITARYRANK CHILD^{-1} MILITARYCOMMAND SERVICESTARTYEAR OFFICE BATTLE SPOUSE^{-1} KNOWNFOR^{-1} PREDECESSOR FOUNDATIONPERSON^{-1} MONARCH^{-1} PREDECESSOR^{-1} ACTIVEYEARSSTARTYEAR ACTIVEYEARSENDYEAR STARRING^{-1} LIEUTENANT PARENT GOVERNOR^{-1} HOMEPAGE RESIDENCE APPOINTER^{-1} ...

Using Change to Simulate Errors

Duboue, Dominguez, Estrella. *On the Robustness of Standalone Referring Expression Generation Algorithms Using RDF Data.* WebNLG 2016.

- Three algorithms of REG on anachronistic input.
  - On old data, produce a referring expression, check whether holds on new data.
- We found poor results with marginal differences among the algorithms.
  - Gardent's algorithm might be ahead but using closed world assumptions.
  - Nice task and problem, worth extending.

WebNLG 2016 results

Algorithm	Execution Errors	Dice	Omission Errors	Inclusion Errors
People – Entity has “birth date”? ⇒ person (3,051 tasks)				
Incremental	232 (5%)	0.48	1,406 (50%)	145 (5%)
Gardent	0 (0%)	0.58	1,089 (36%)	554 (18%)
Graph	15 (0%)	0.38	1,870 (62%)	20 (0%)
Organizations – Entity has “creation date”? ⇒ organization (2,370 tasks)				
Incremental	1,386 (45%)	0.69	305 (31%)	3 (0%)
Gardent	829 (27%)	0.70	338 (22%)	357 (23%)
Graph	934 (31%)	0.06	1,347 (94%)	2 (0%)

- Alusivo: Open Source implementation of REG algos
  - (MPL) `https://github.com/DrDub/Alusivo`
  - Java, Maven, RDF-based
  - CSP-based algorithms, Graph isomorfism-based algorithms, etc

Using Robustness to Learn the Property Ordering

Duboue, Dominguez. *Using Robustness to Learn to Order Semantic Properties in Referring Expression Generation* Iberamia 2016.

- We tried to learn the PO using errors on generated referring expressions as a metric
- Intuitions
  - A good referring expression should refer to stable properties
- Results
  - Robustness helps to learn ordernings
  - But popularity on DBpedia is a stronger signal

Iberamia Metrics

- Measure learned POs against the hand-written PO
  - Kendall's τ [Lebanon and Lafferty, 2002]:

$$\tau = 1 - \frac{2(\text{number of inversion})}{N(N-1)/2}$$

    - too strict, moved to a metric that considers the REs being generated rather than the exact ordering
  - Dice over selected properties.
    - Seemingly very different POs produce comparable results
    - Metric of choice
- Do observable variables change similarly to target?
  - Spearman's rho

First Experiments

- Experiment: Correlations over People

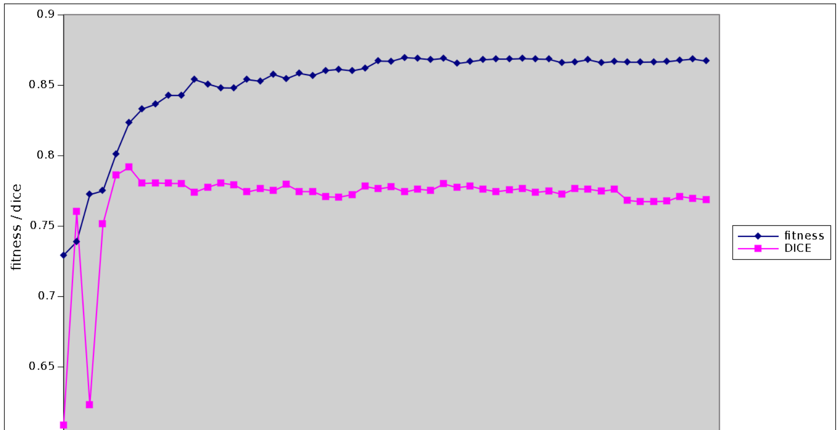
Exp/metric	length	Dice	exclusion errors	inclusion errors
Hand-written	-0.018	-0.215	0.185	0.397
Popularity	-0.226	0.232	-0.258	0.394

- Experiment: overfit fitness function
  - A function that approximates the Dice for the hand-picked PO using the observable variables. Linear regression:

    $$target = -1.608 * length + 15.5279 * inclusion + 0.8787 * exclusion + 1.9403$$

  - Pearson's correlation coefficient of 0.872

- Experiment:
  Can the GA learn?



Main Experiments

- Experiment: Genetic Algorithm over organizations using fitness trained on peopled
  - Disappointment: after 50 generations, we get a Dice property overlap of only 0.435
  - When popularity PO achieves 0.93.
- This is our main negative result

Closing Experiments

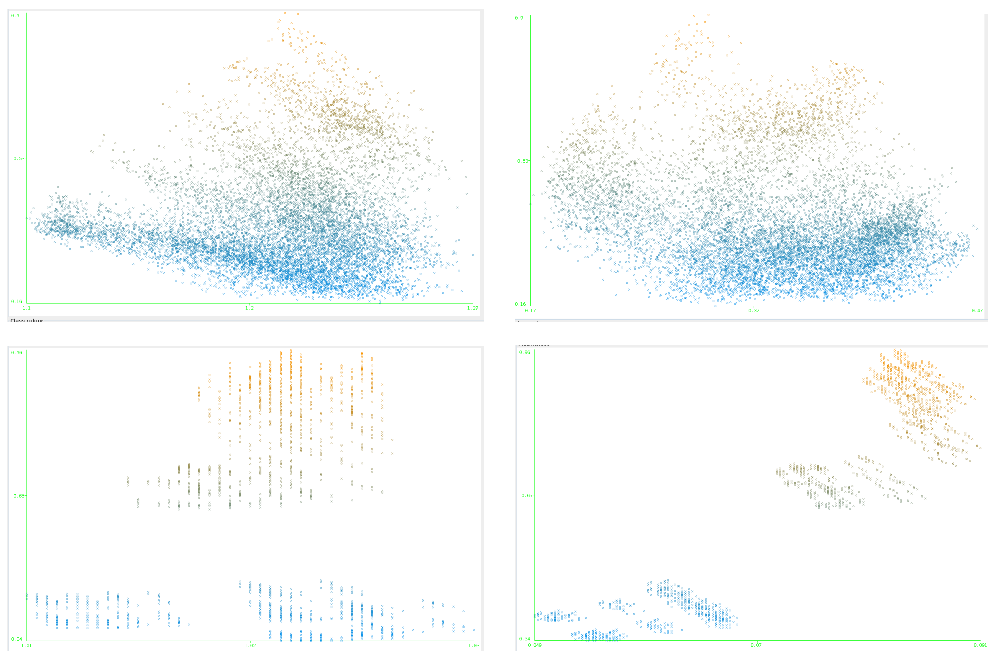
- Experiment: Correlations over Organizations
  - Strong Spearman's rho, but very different from people's numbers.

Exp/metric	length	Dice	exclusion errors	inclusion errors
Hand-written	0.059	0.832	-0.834	0.840
Popularity	-0.064	0.864	-0.866	0.841

- Experiment: Only length & inclusion errors
  - Preliminary result, insight obtained from looking at both tables (test set)
  - Trained on people, Dice on organizations of 0.906
  - Trained on organizations, Dice on people of 0.608
  - Below the popularity PO but more generalization strength
- Experiment: GA using only inclusion errors
  - Dice of 0.272 (people) and 0.361 (organizations)
  - Robustness alone is not enough, combining it with length is key.

Learning PO Discussion

- Main result: correlation between hand-written PO and robustness
- Lack of generalization: organizations change differently from people (hypothesis)



Other Work

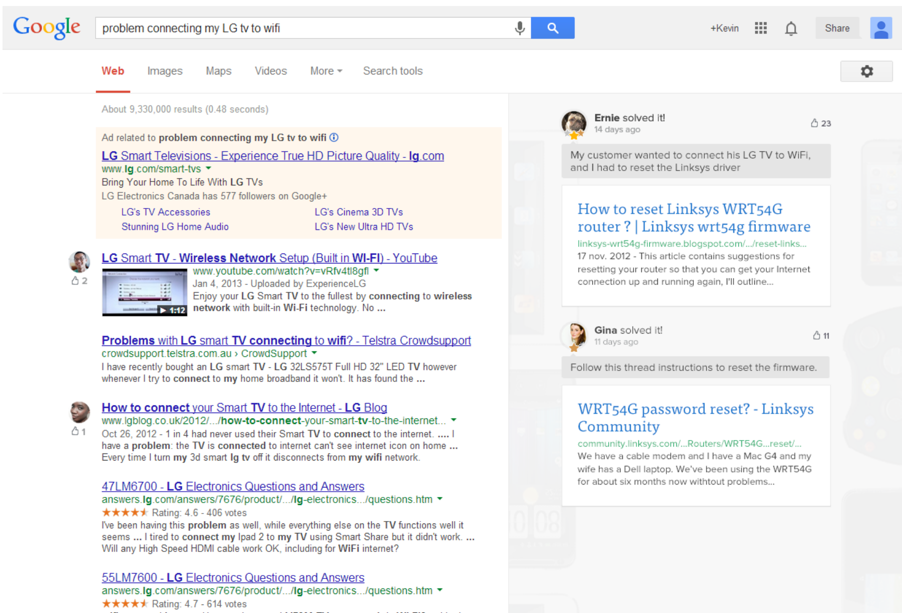
- Radialpoint Reveal / Canadian AI
- Thoughtland / EWNLG
- Hybrid IE Systems / IE4OpenData

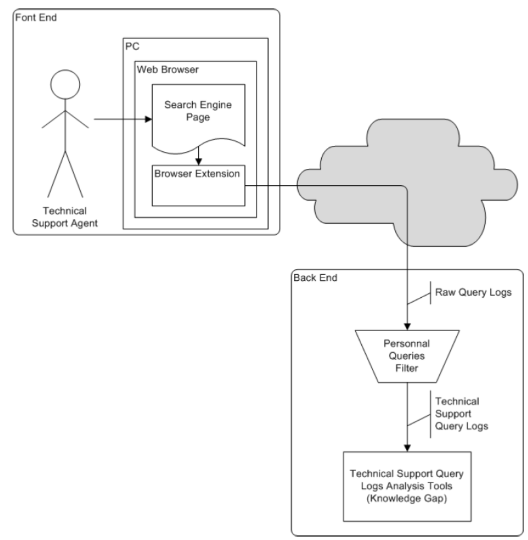
Radialpoint Reveal

- In Montreal I was part of a multi-year effort to build an enhanced search engine experience for tech support agents: Radialpoint Reveal
- We worked on multiple fronts, including
  - Better search by pooling results across agents
  - Identifying tech support rich pages vs. other device-related pages
  - Identifying solutions to problems within a page and matching problem statements to search terms, devices and models

Canadian AI

Neto, Desaulniers, Duboue, Smirnov. *Filtering Personal Queries from Mixed-Use Query Logs*. (Best Paper Award) Canadian AI 2014.

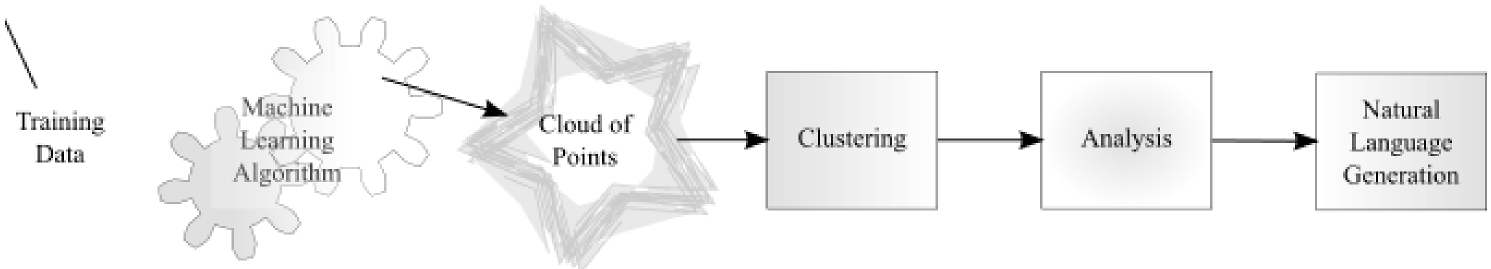




- Similar to Google Trends but for work searches
- We filter out 78.7% of private queries losing only 9.3% of the business queries
  - Kappa 0.87 for annotating them

Thoughtland / EWNLG

Duboue. *Thoughtland: Natural Language Descriptions for Machine Learning n-dimensional Error Functions.* ENLG 2013.



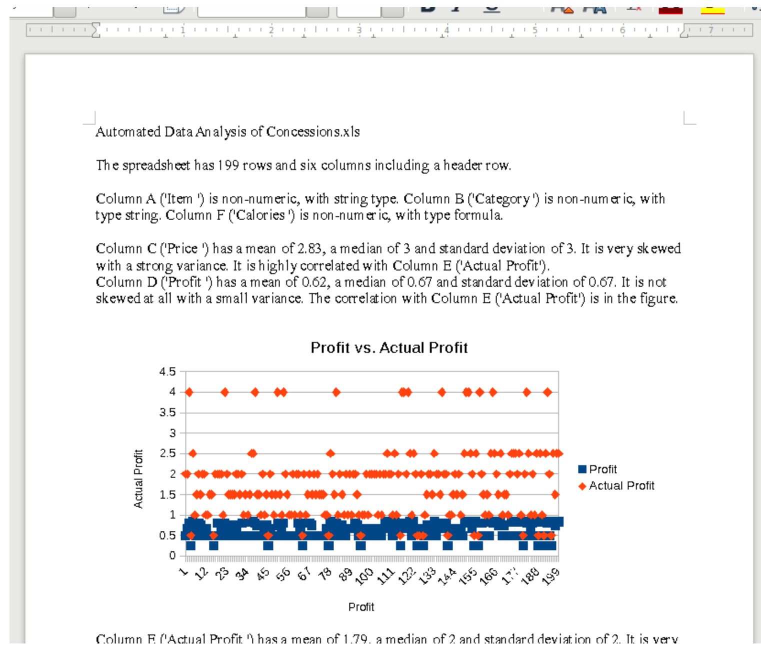
- Cross-validation: cloud of error points
- Cluster with mixture of Dirichlet models: n-balls
- Determine overall size, density, distances to others
- Source code (AGPL): `https://github.com/DrDub/thoughtland`

There are six components and three dimensions. Component One is big, components Two, Three and Four are small and component Five is giant. Component Five is sparse and components Two, Three and Four are very dense. Components One and Two are at a good distance from each other. The rest are all far from each other.

Textualization: Data Report

Duboue. *Automatic Reports from Spreadsheets: Data Analysis for the Rest of Us.* (Demo) INLG 2016.

	A	B	C	D	E	F
1	Item	Category	Price	Profit	Actual Profit	Calories
2	Beer	Beverages	$ 4.00	50%	$ 2.00	200
3	Hamburger	Hot Food	$ 3.00	67%	$ 2.00	320
4	Popcorn	Hot Food	$ 5.00	80%	$ 4.00	500
5	Pizza	Hot Food	$ 2.00	25%	$ 0.50	480
6	Bottled Water	Beverages	$ 3.00	83%	$ 2.50	0
7	Hot Dog	Hot Food	$ 1.50	67%	$ 1.00	265
8	Chocolate Dipped Cone	Frozen Treats	$ 3.00	50%	$ 1.50	300
9	Soda	Beverages	$ 2.50	80%	$ 2.00	120
10	Chocolate Bar	Candy	$ 2.00	75%	$ 1.50	255
11	Hamburger	Hot Food	$ 3.00	67%	$ 2.00	320
12	Beer	Beverages	$ 4.00	50%	$ 2.00	200
13	Hot Dog	Hot Food	$ 1.50	67%	$ 1.00	265
14	Licorice Rope	Candy	$ 2.00	50%	$ 1.00	280
15	Chocolate Dipped Cone	Frozen Treats	$ 3.00	50%	$ 1.50	300
16	Nachos	Hot Food	$ 3.00	50%	$ 1.50	560
17	Pizza	Hot Food	$ 2.00	25%	$ 0.50	480
18	Beer	Beverages	$ 4.00	50%	$ 2.00	200



- Domain independent tabular data verbalization
- Conduit for domain-dependent customization through consulting engagements

Hybrid IE Systems

- Last summer I taught a 15hs winter school on Hybrid IE Systems.
  - In Universidad de Buenos Aires
  - Generalities of IE systems
  - Rule-based systems using Apache RuTA
  - CRFs using Mallet
  - Underlining framework Apache UIMA
- Slides: (in English, CC-BY-SA)

  `https://github.com/IE4OpenData/ECI2016T2`
- Code: (Apache Licenced)

  `https://github.com/IE4OpenData/Octroy`

IE 4 OpenData

- The course spawned a project on using Information Extraction over Open Data
  - Better transparency in democracy
- `http://ie4opendata.org`

Conclusions

- DBpedia/Wikinews is a suitable source for doing research on robust REG algorithms.
- DBpedia is fine as a one time NLG resource
  - Usage over time requires better algorithms
- Links
  - `http://duboue.net`
  - `https://twitter.com/pabloduboue`
  - `https://github.com/DrDub`
  - `https://scholar.google.com/citations?user=Exngg_MAAAAJ&hl=en`

References

[Bizer et al., 2009] Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., and Hellmann, S. (2009). DBpedia-a crystallization point for the web of data. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7(3):154–165.

[Dale and Reiter, 1995] Dale, R. and Reiter, E. (1995). Computational interpretations of the gricean maxims in the generation of referring expressions. *Cognitive Science*, 19(2):233–263.

[Lebanon and Lafferty, 2002] Lebanon, G. and Lafferty, J. (2002). Combining rankings using conditional probability models on permutations. In Sammut, C. and A. Hoffmann, e., editors, *Proceedings of the 19th International Conference on Machine Learning*, San Francisco, CA. Morgan Kaufmann Publishers.

[McCullagh and Yang, 2008] McCullagh, P. and Yang, J. (2008). How many clusters? *Bayesian Analysis*, 3(1):101–120.

[Pacheco et al., 2012] Pacheco, F., Duboue, P. A., and Domínguez, M. A. (2012). On the feasibility of open domain referring expression generation using large scale folksonomies. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, NAACL HLT '12, pages 641–645, Stroudsburg, PA, USA. Association for Computational Linguistics.

[Siddharthan et al., 2011] Siddharthan, A., Nenkova, A., and McKeown, K. (2011). Information status distinctions and referring expressions: An empirical study of references to people in news summaries. *Computational Linguistics*, 37(4):811–842.