Modeling Topic Hierarchies with the Recursive Chinese Restaurant Process

Joon Hee Kim, Dongwoo Kim, Suin Kim, Alice Oh
KAIST

Outline of this presentation

- Introduce **topic modeling**
  - with illustration of **LDA**
- Explain the importance of **hierarchy** in data mining
  - and explain **rCRP**
  - and show **two new evaluation metrics**

Topic Models

- Statistical models for discovering the abstract "**topics**" that occur in a collection of documents
- Input
  - Collection of documents
- Output
  - **Word distribution** of each topic
  - **Topic distribution** of each document
- Line of researches: LSI, PLSI, LDA, HDP

http://www.nytimes.com/2012/10/21/sports/how-armstrongs-wall-fell-one-rider-at-a-time.html

October 20, 2012

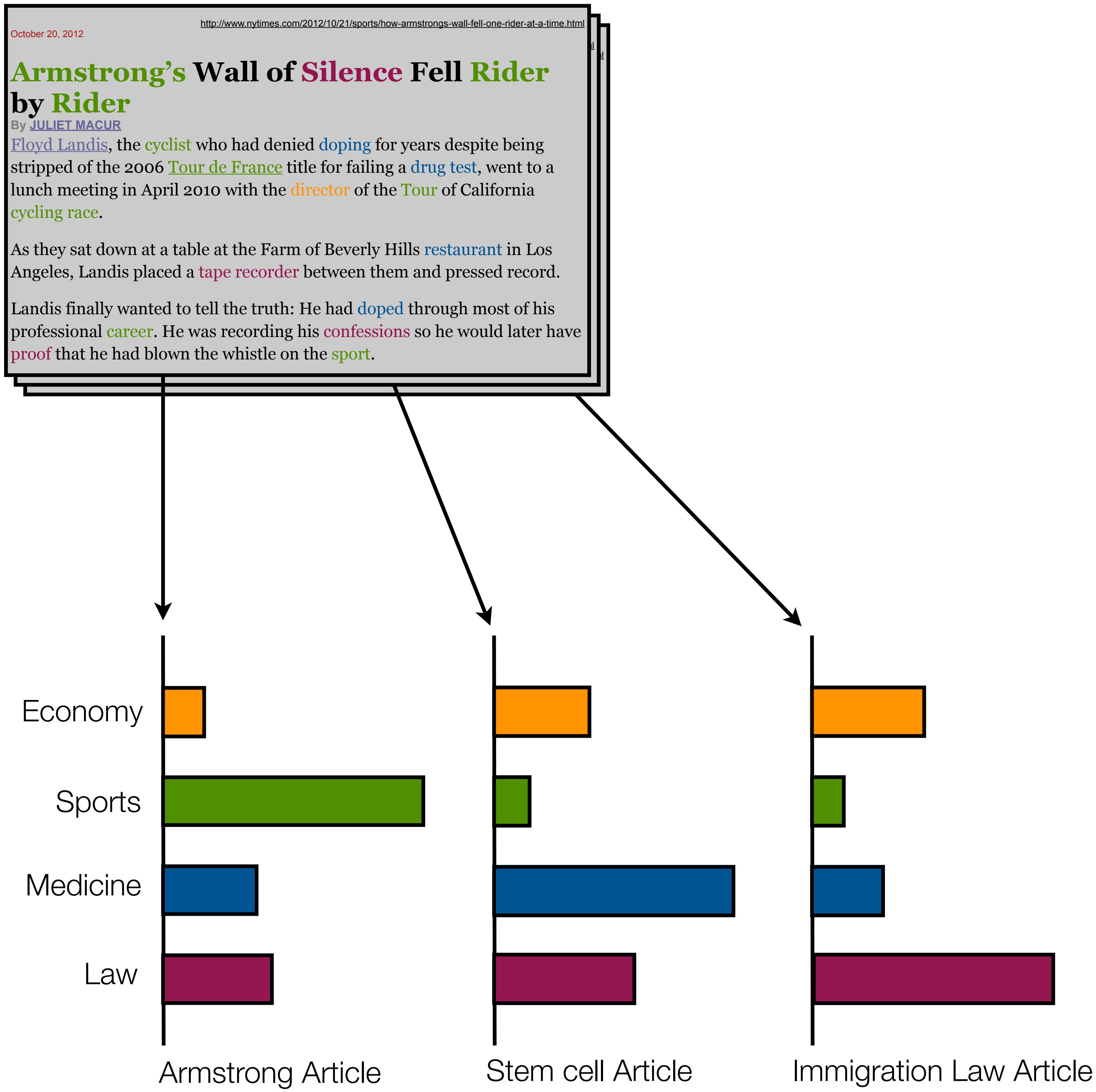
Armstrong's Wall of Silence Fell Rider by Rider

By JULIET MACUR

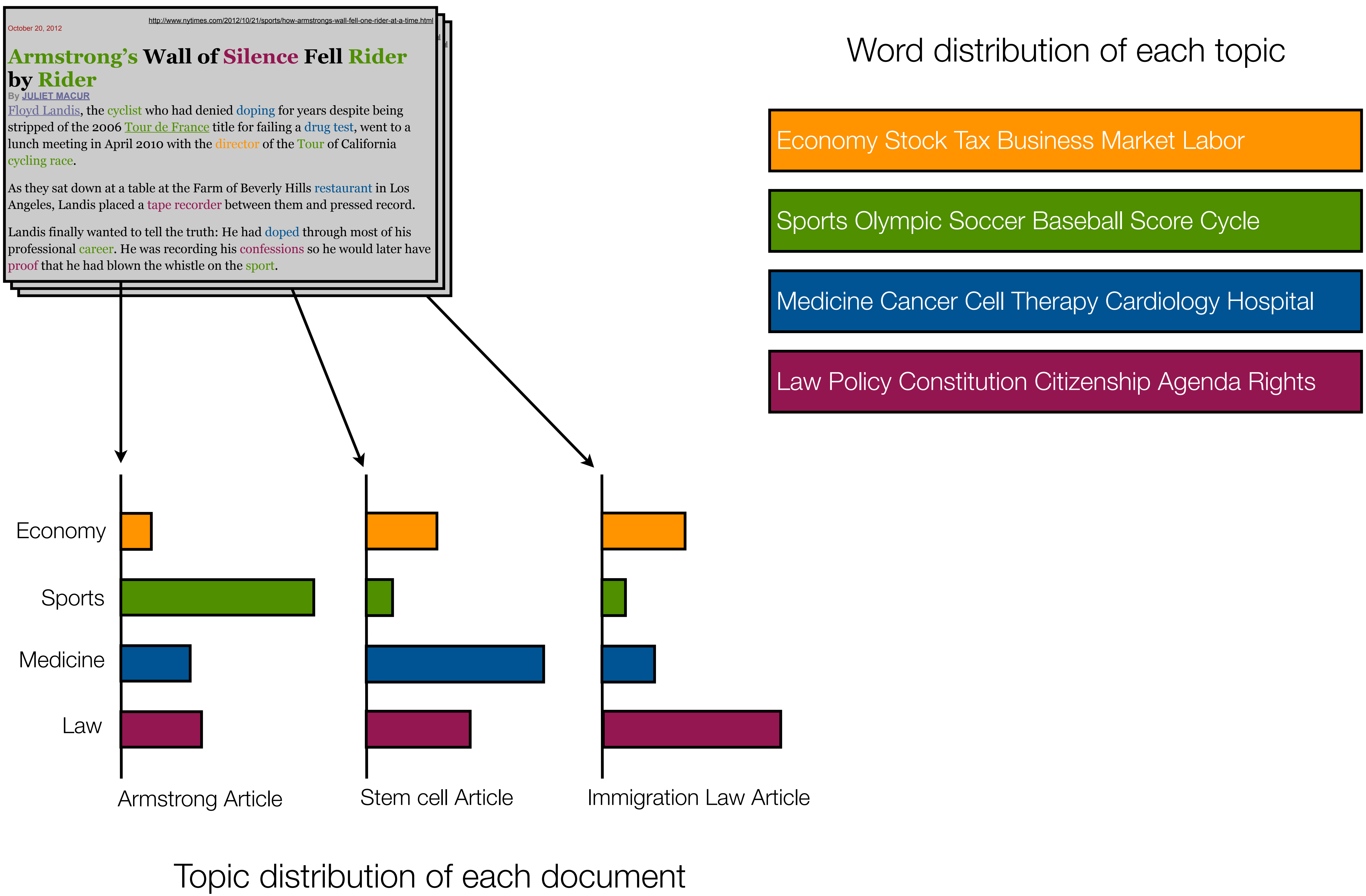
Floyd Landis, the cyclist who had denied doping for years despite being stripped of the 2006 Tour de France title for failing a drug test, went to a lunch meeting in April 2010 with the director of the Tour of California cycling race.

As they sat down at a table at the Farm of Beverly Hills restaurant in Los Angeles, Landis placed a tape recorder between them and pressed record.

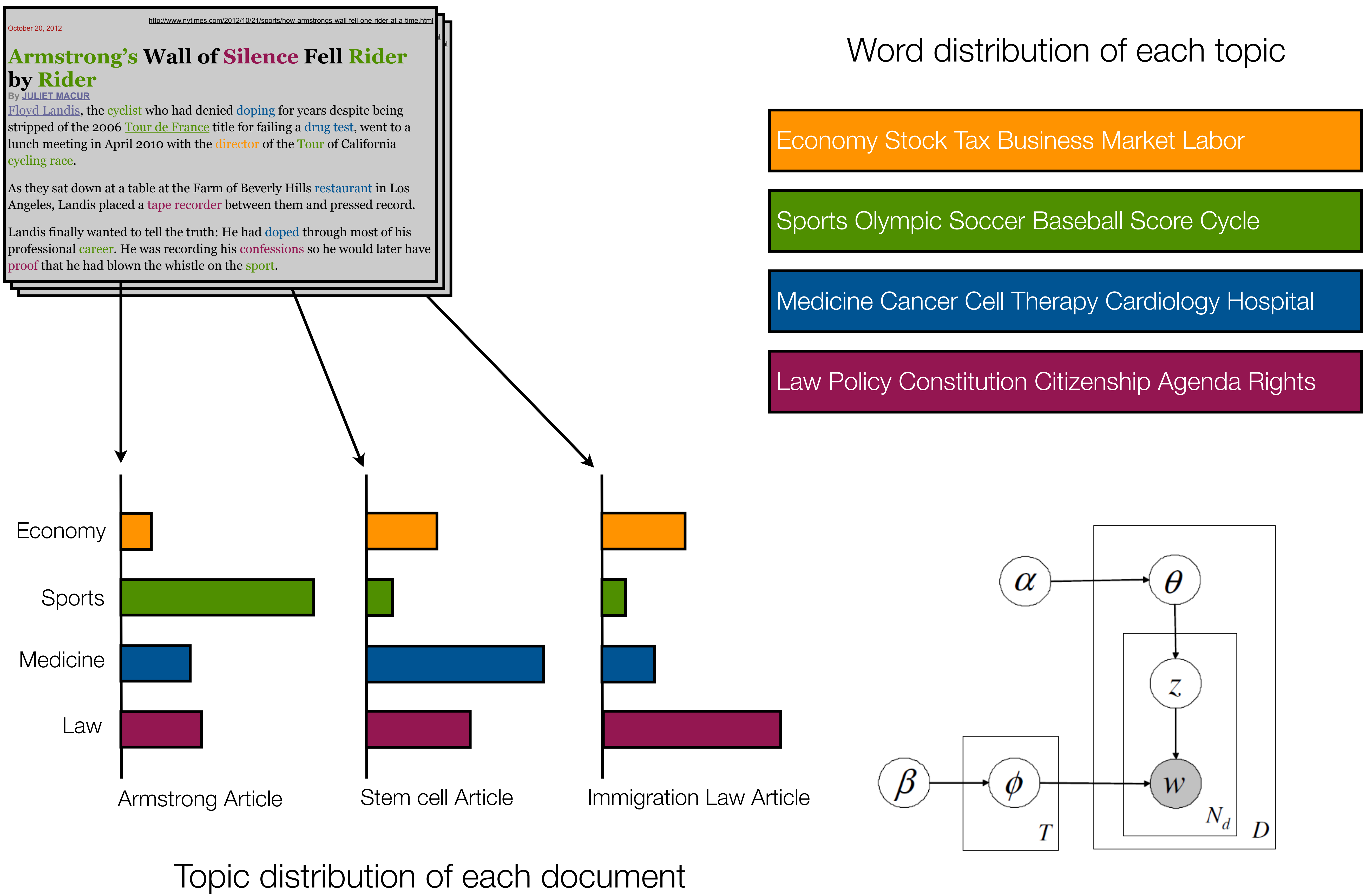
Landis finally wanted to tell the truth: He had doped through most of his professional career. He was recording his confessions so he would later have proof that he had blown the whistle on the sport.



Topic distribution of each document



October 20, 2012
http://www.nytimes.com/2012/10/21/sports/how-armstrongs-wall-fell-one-rider-at-a-time.html
Armstrong's Wall of Silence Fell Rider by Rider
By JULIET MACUR
Floyd Landis, the cyclist who had denied doping for years despite being stripped of the 2006 Tour de France title for failing a drug test, went to a lunch meeting in April 2010 with the director of the Tour of California cycling race.
As they sat down at a table at the Farm of Beverly Hills restaurant in Los Angeles, Landis placed a tape recorder between them and pressed record.
Landis finally wanted to tell the truth: He had doped through most of his professional career. He was recording his confessions so he would later have proof that he had blown the whistle on the sport.
Word distribution of each topic
Economy Stock Tax Business Market Labor
Sports Olympic Soccer Baseball Score Cycle
Medicine Cancer Cell Therapy Cardiology Hospital
Law Policy Constitution Citizenship Agenda Rights
Economy
Sports
Medicine
Law
Armstrong Article
Stem cell Article
Immigration Law Article
Topic distribution of each document


October 20, 2012
http://www.nytimes.com/2012/10/21/sports/how-armstrongs-wall-fell-one-rider-at-a-time.html
Armstrong's Wall of Silence Fell Rider by Rider
By JULIET MACUR
Floyd Landis, the cyclist who had denied doping for years despite being stripped of the 2006 Tour de France title for failing a drug test, went to a lunch meeting in April 2010 with the director of the Tour of California cycling race.
As they sat down at a table at the Farm of Beverly Hills restaurant in Los Angeles, Landis placed a tape recorder between them and pressed record.
Landis finally wanted to tell the truth: He had doped through most of his professional career. He was recording his confessions so he would later have proof that he had blown the whistle on the sport.
Word distribution of each topic
Economy Stock Tax Business Market Labor
Sports Olympic Soccer Baseball Score Cycle
Medicine Cancer Cell Therapy Cardiology Hospital
Law Policy Constitution Citizenship Agenda Rights
Economy
Sports
Medicine
Law
Armstrong Article
Stem cell Article
Immigration Law Article
Topic distribution of each document
α
θ
z
w
β
φ
T
N_d
D

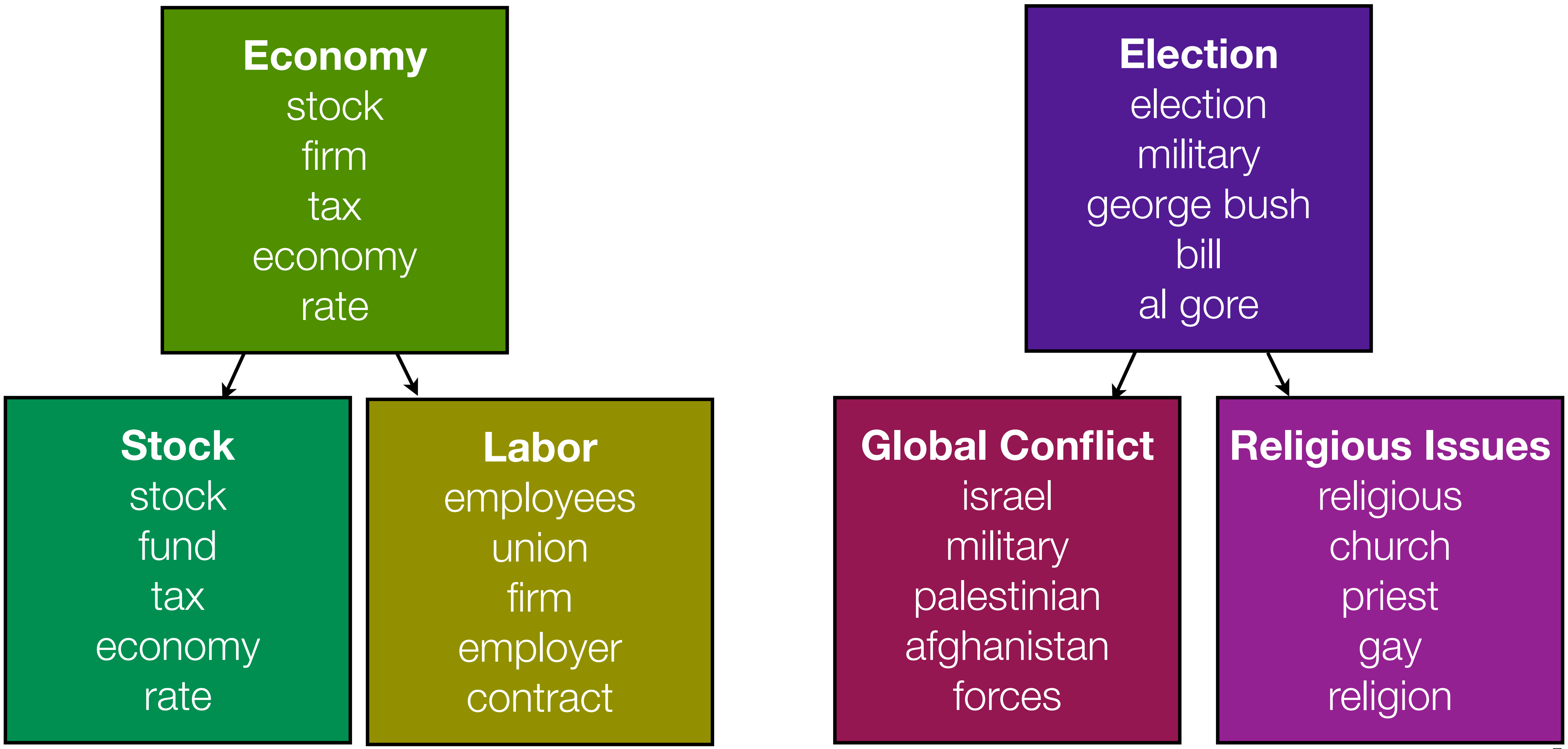
Problem with LDA

- No relation / structure among the retrieved topics

Stock stock fund tax economy rate	**Global Conflict** israel military palestinian afghanistan forces	**Economy** stock firm tax economy rate
Labor employees union firm employer contract	**Election** election military george bush bill al gore	**Religious Issues** religious church priest gay religion

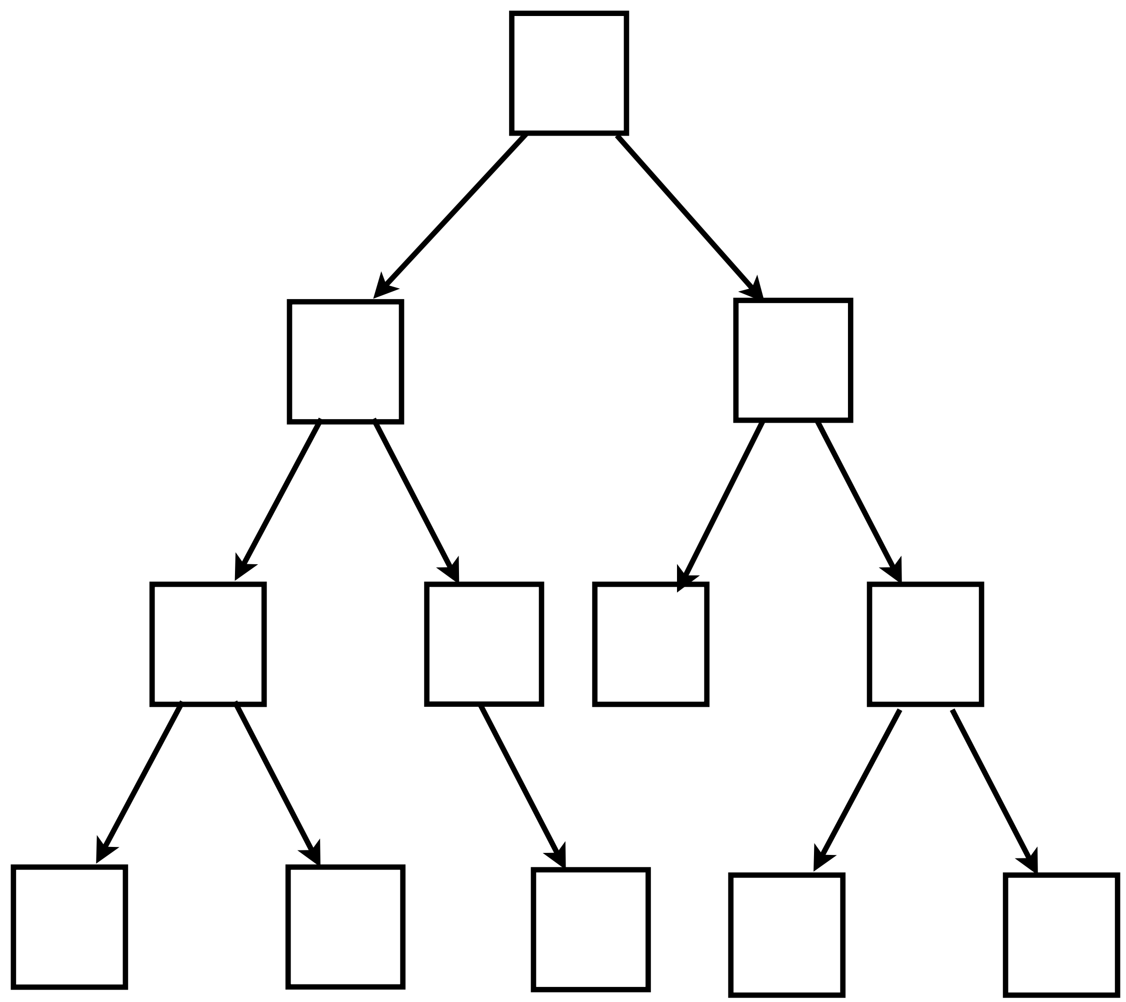
Problem with LDA

- No relation / structure among the retrieved topics



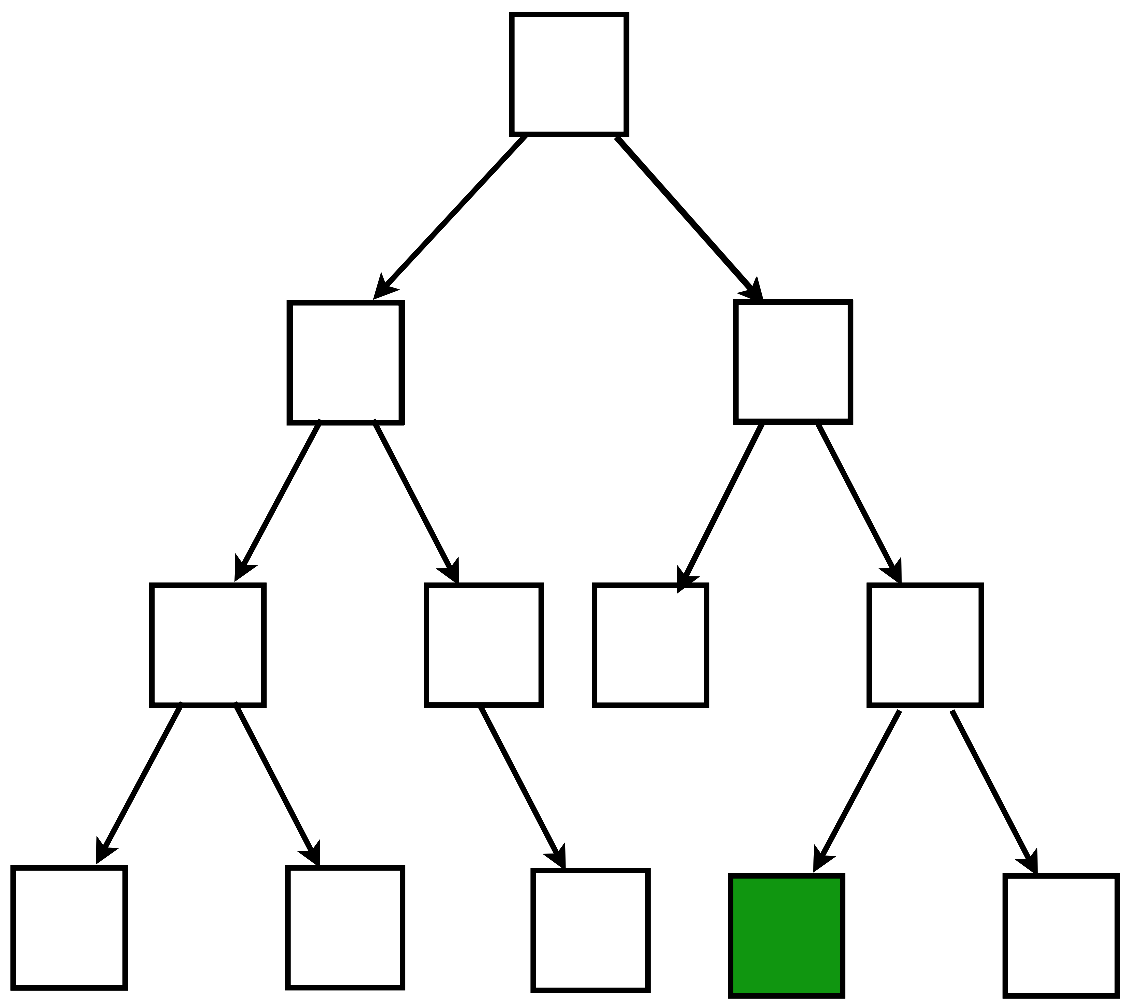
Related Works

- Each model has **unique assumptions** about how documents are generated from the topic tree



Related Works

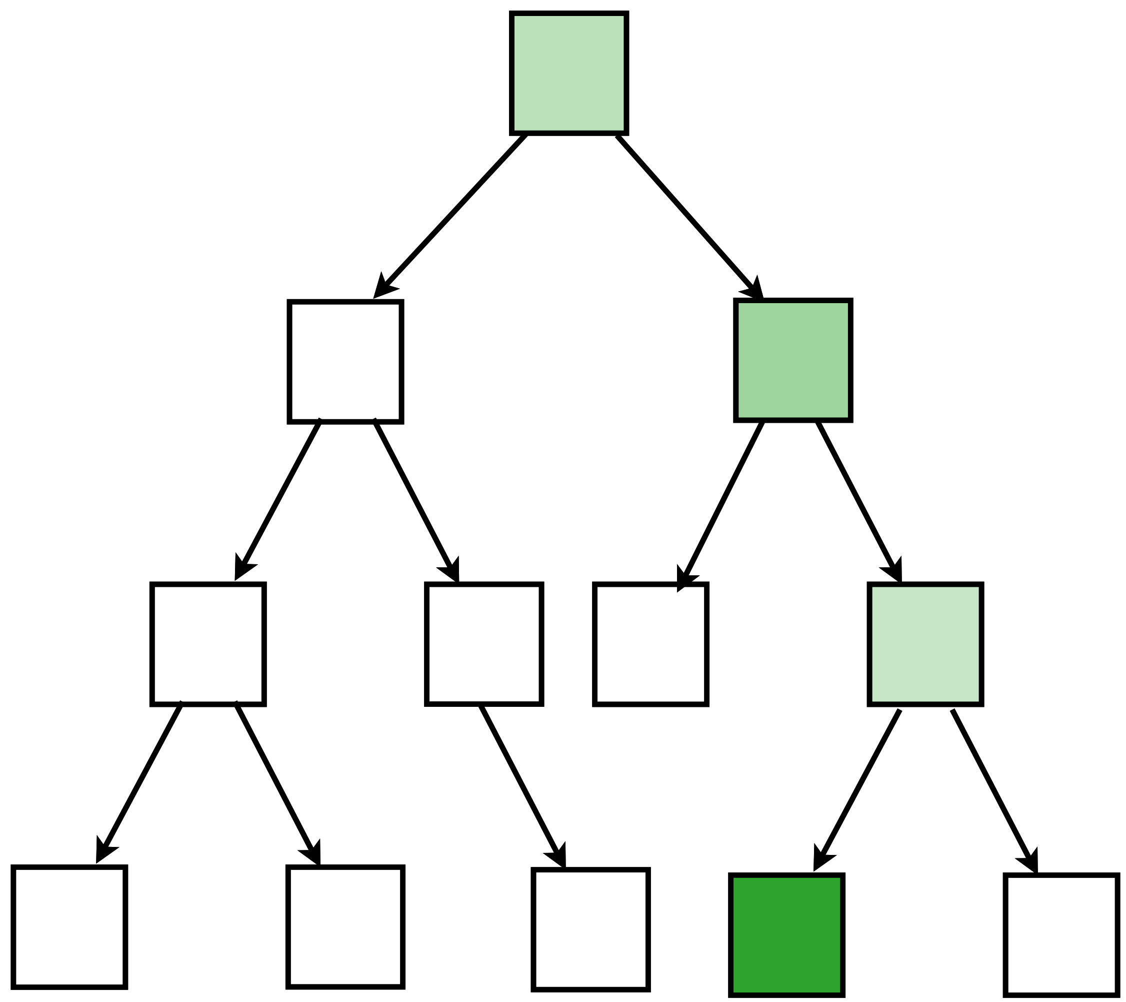
- TS-SB (Tree-structured Stick-breaking)
  Document is generated by **a single topic** in the topic tree



Adams, Ryan Prescott, Zoubin Ghahramani, and Michael I. Jordan. "Tree-structured stick breaking for hierarchical data." *Advances in Neural Information Processing Systems* 23 (2010): 19-27.

Related Works

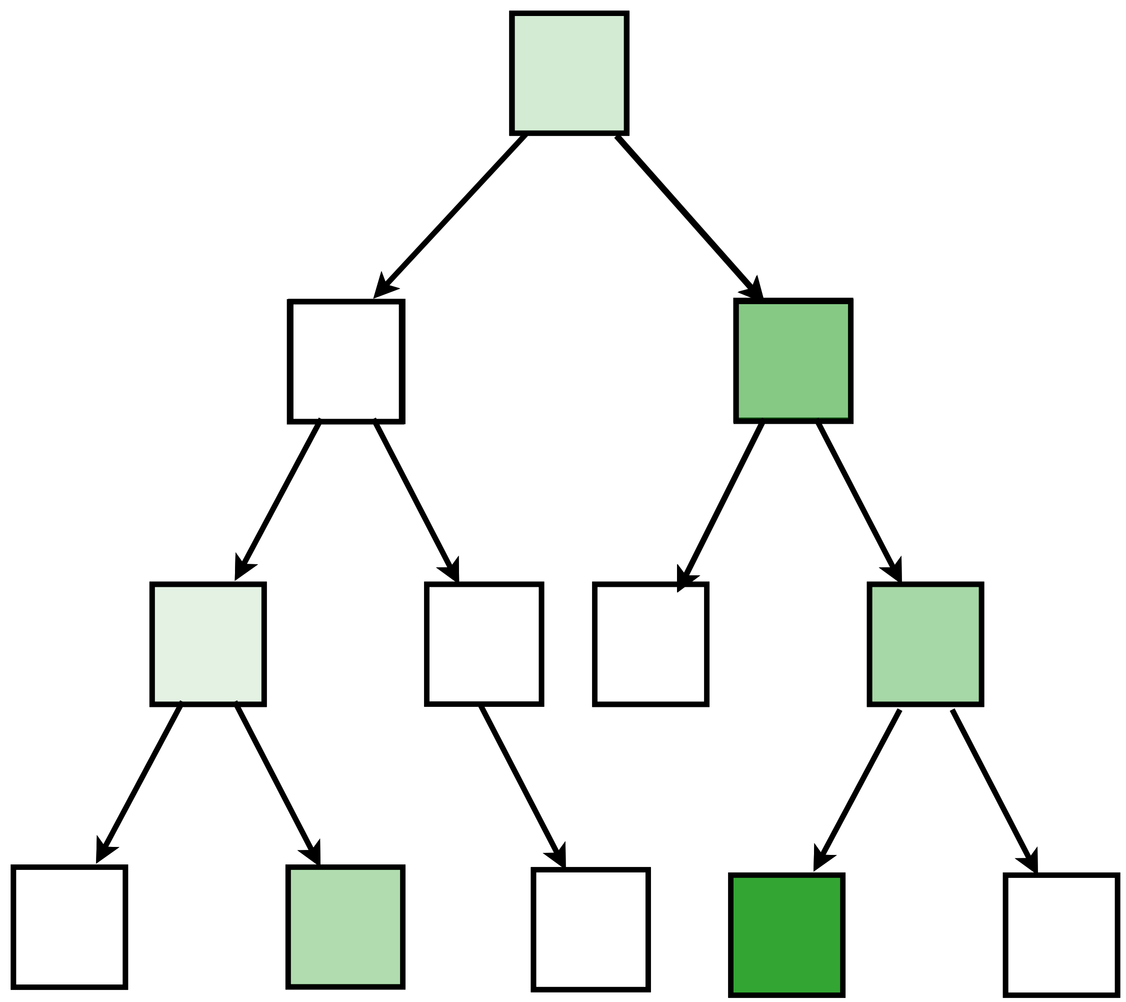
- nCRP (nested Chinese Restaurant Process): A document is generated by a **topic path** from the topic tree



Griffiths, David M. Blei Thomas L., and Michael I. Jordan Joshua B. Tenenbaum. "Hierarchical topic models and the nested Chinese restaurant process." *Advances in Neural Information Processing Systems 16: Proceedings of the 2003 Conference*. Vol. 16. MIT Press, 2004.

Related Works

- rCPR (recursive Chinese Restaurant Process): A document is generated by **the entire topic tree**



Our Model: Overview

- Our model extends HDP (Hierarchical Dirichlet Process), which utilizes the **Chinese Restaurant Franchise** metaphor
- Metaphor:
    - As customers enter a restaurant, they are assigned a table.
    - For each table, a dish is served from the global menu

Our Model: Overview

- Our model extends HDP (Hierarchical Dirichlet Process), which utilizes the **Chinese Restaurant Franchise** metaphor
- Metaphor:
  - As customers enter a restaurant, they are assigned a table.
  - For each table, a dish is served from the global menu
    - Customer = Word
    - Table = Group of words in a document
    - Restaurant = Document
    - (HDP) Global menu = Global topics
    - **(rCRP) Global menu tree = Global topic tree**

Our Model: Overview

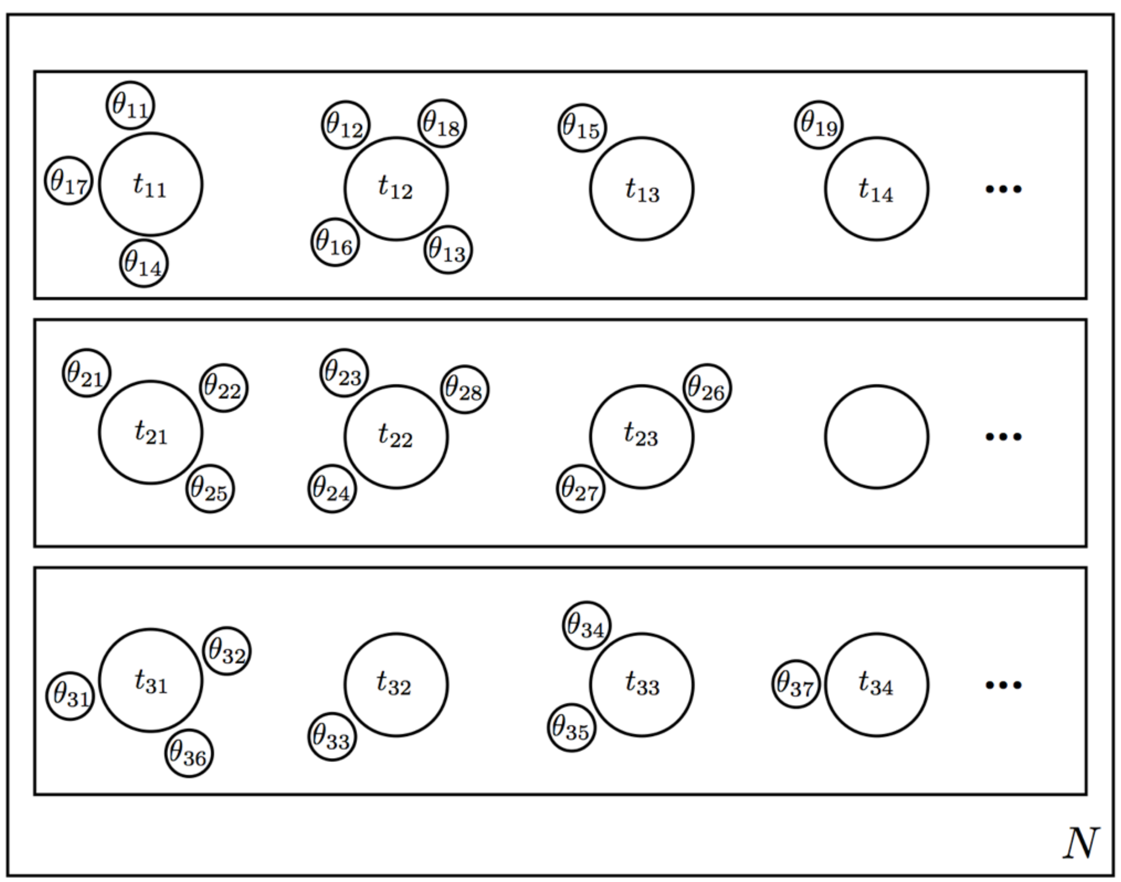
- Our model extends HDP (Hierarchical Dirichlet Process), which utilizes the **Chinese Restaurant Franchise** metaphor
- Metaphor:
  - As customers enter a restaurant, they are assigned a table.
  - For each table, a dish is served from the global menu
    - Customer = Word
    - Table = Group of words in a document
    - Restaurant = Document
    - (HDP) Global menu = Global topics
    - **(rCRP) Global menu tree = Global topic tree**
- Words in documents are partitioned into groups.
- A topic is assigned for each group of words from the global topics

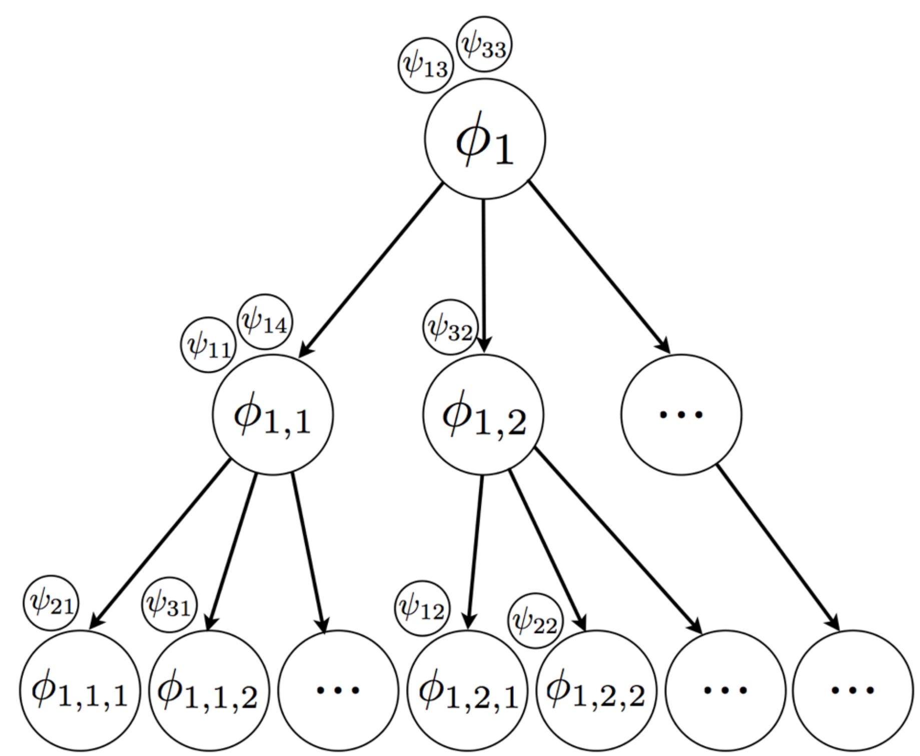
Our Model: Schematic View

- The model employs two levels of assignment processes

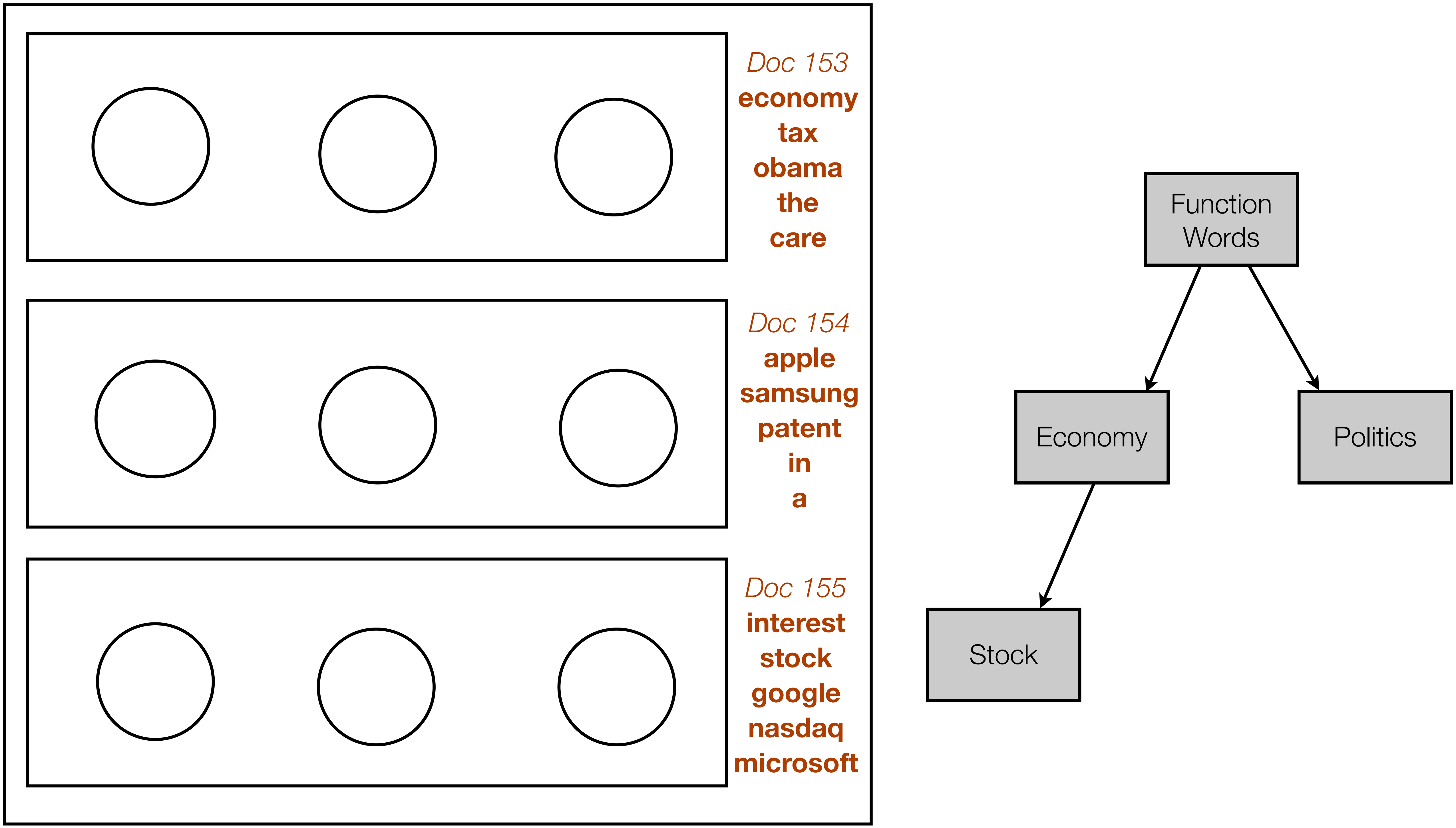
- Second level CRP assigns a table for each customer in a restaurant (= partitions words in documents into groups)

- First level rCRP assigns a dish for each table from the menu tree (= assigns a topic for each group by recursively searching the global menu tree)





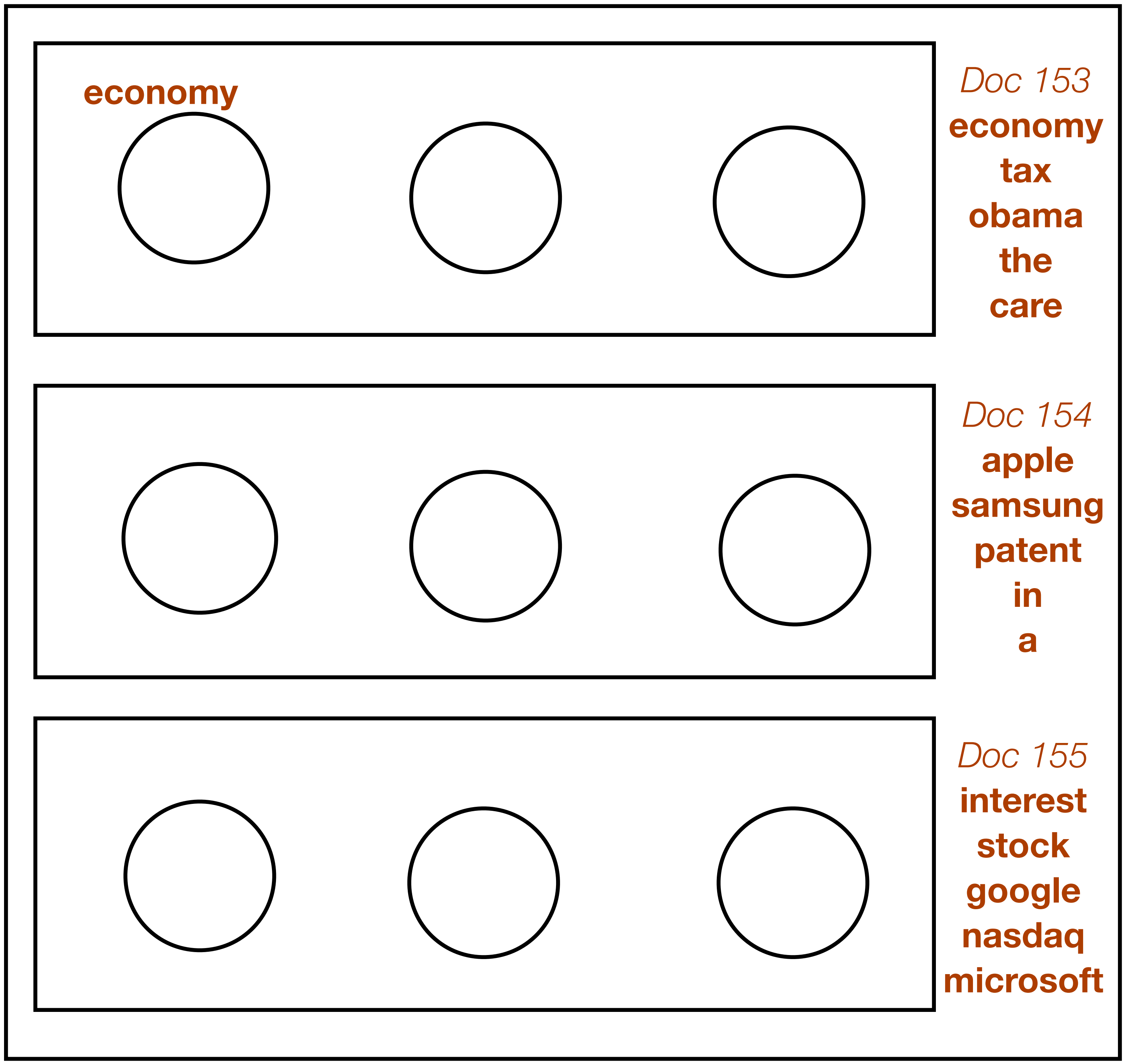
Our Model: Assignment Example



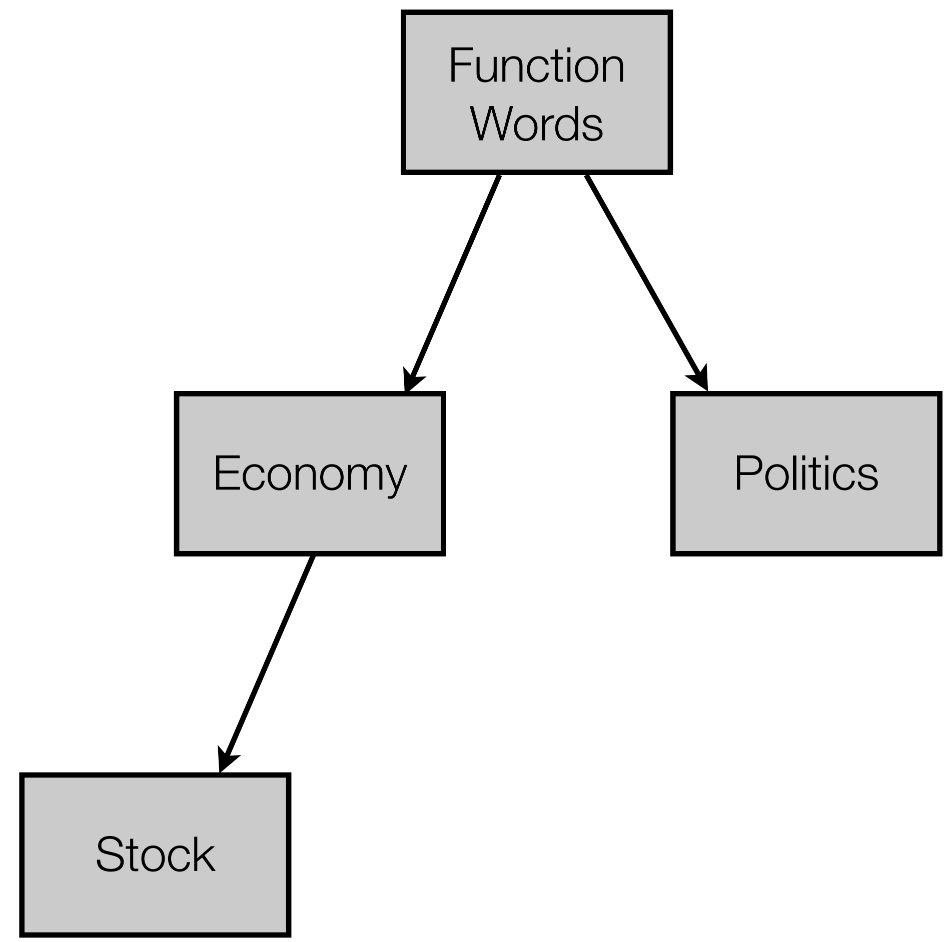
Second level CRP

First level rCRP

Our Model: Assignment Example

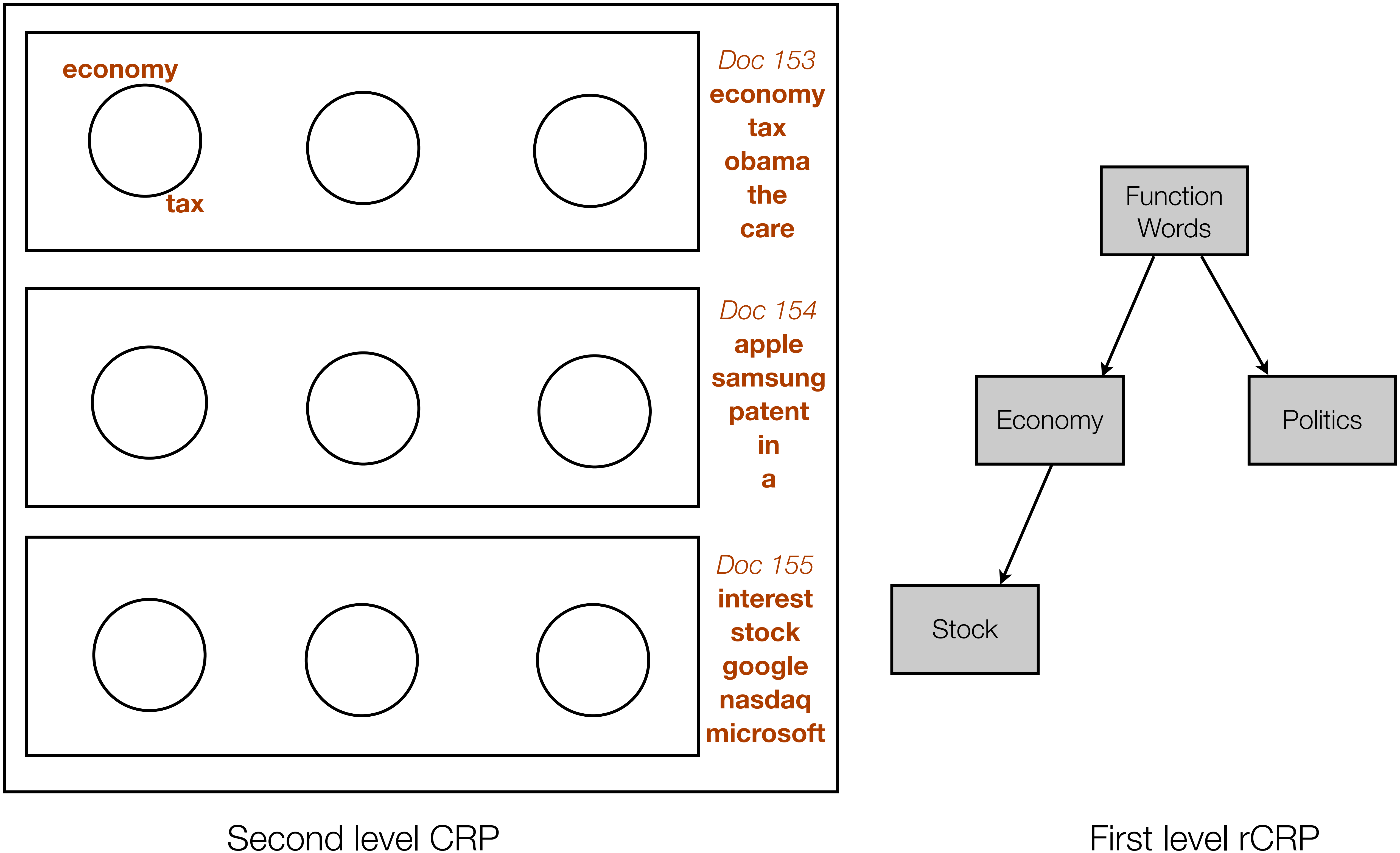


Second level CRP

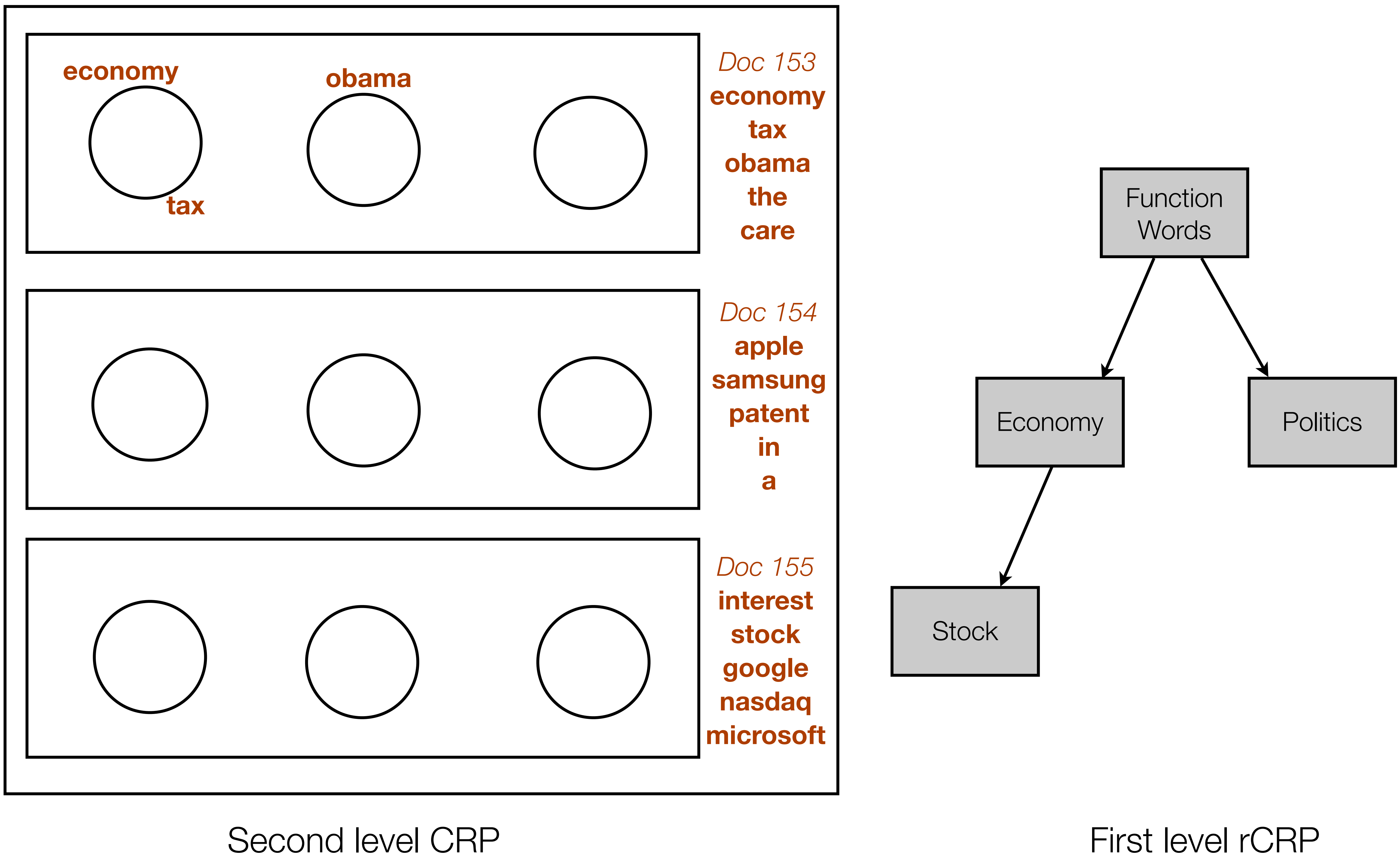


First level rCRP

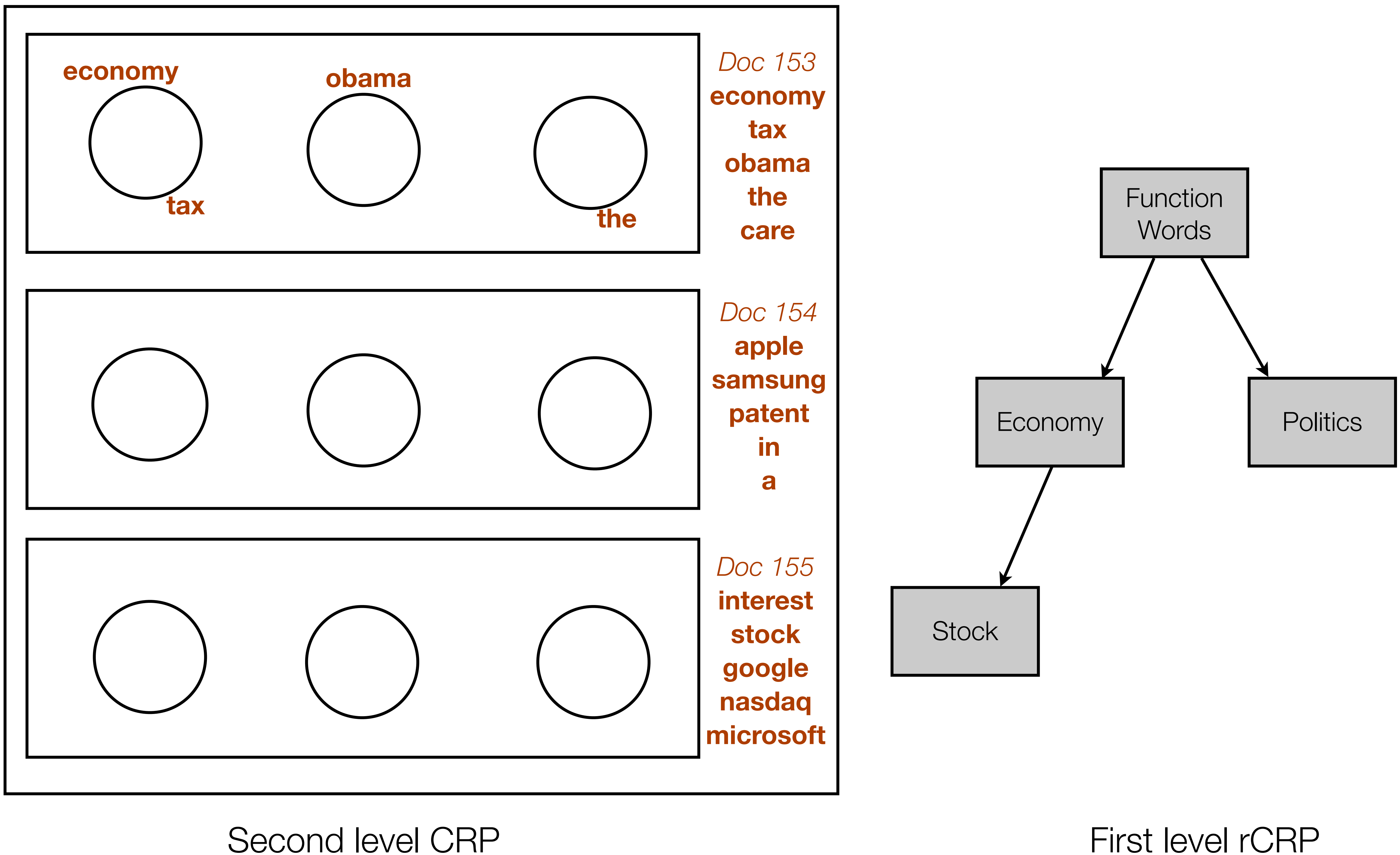
Our Model: Assignment Example



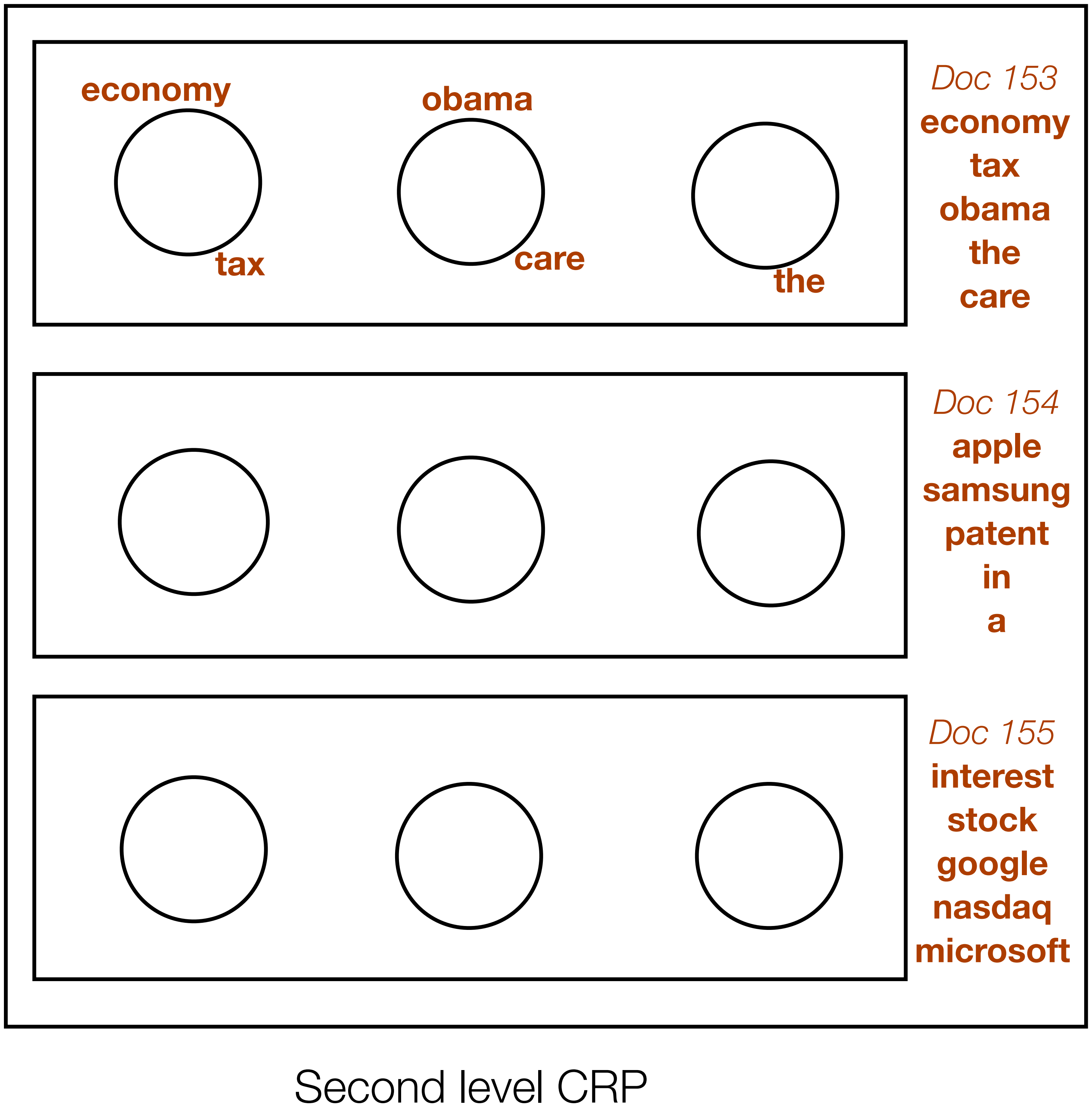
Our Model: Assignment Example

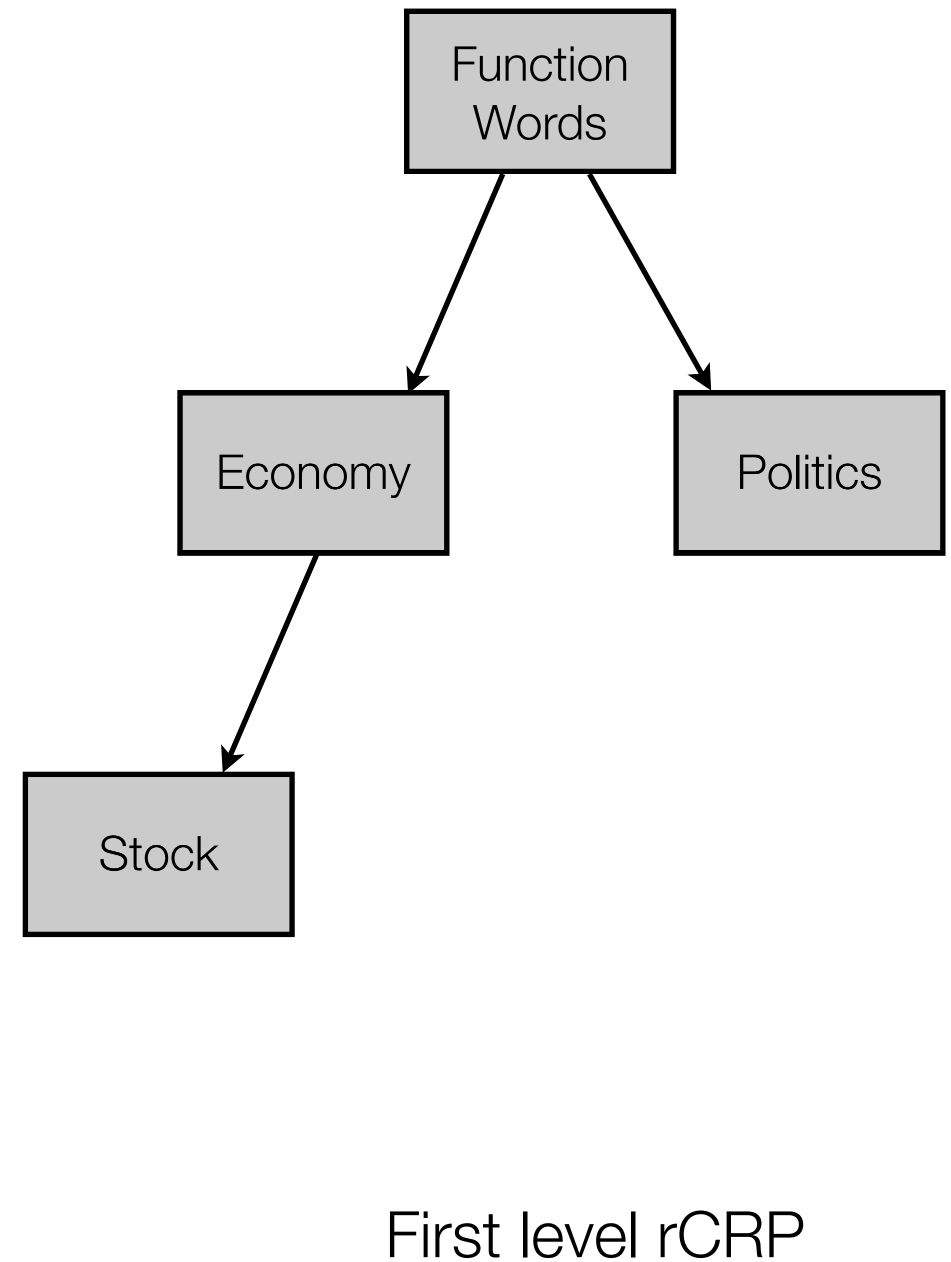


Our Model: Assignment Example

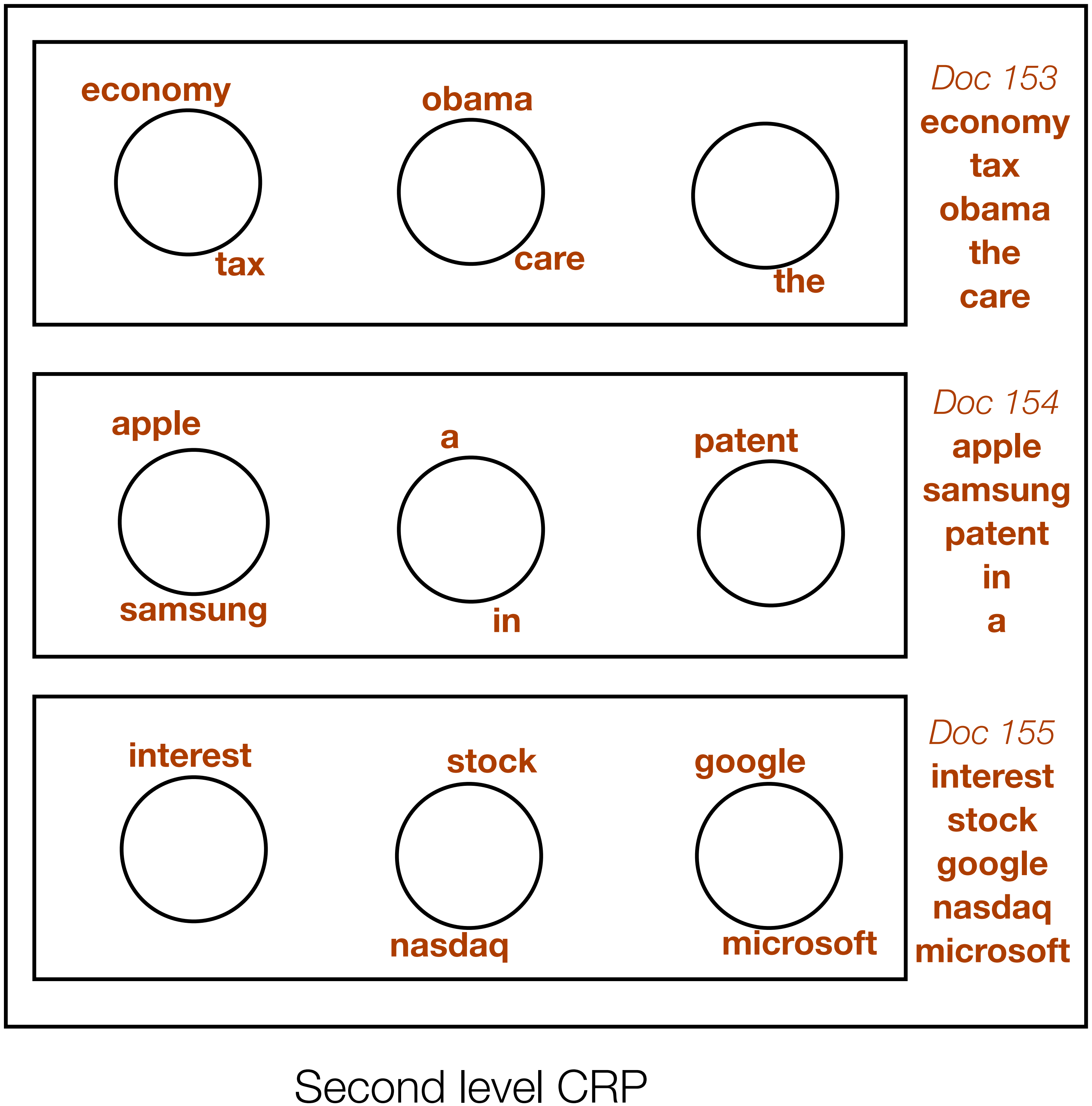


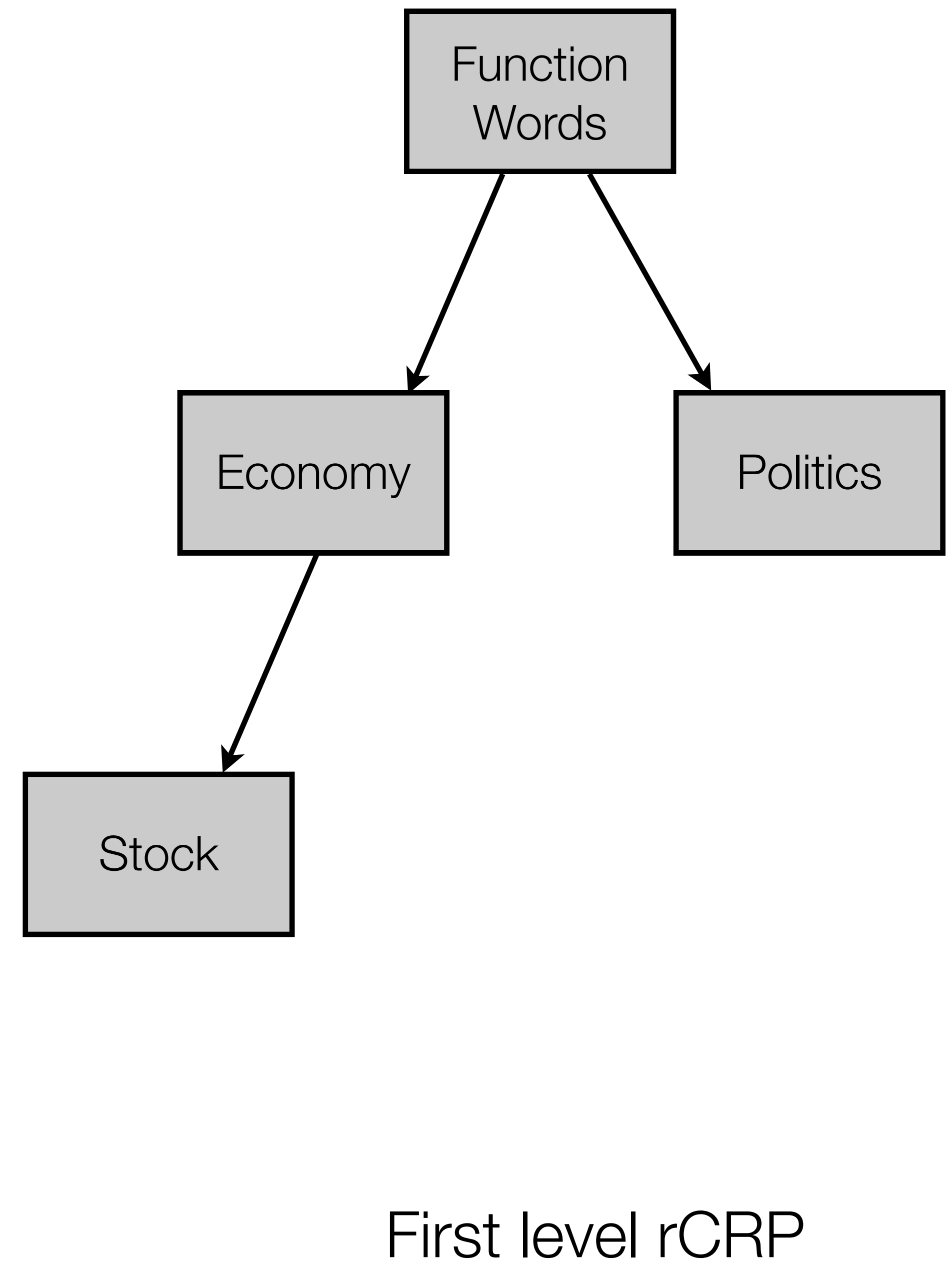
Our Model: Assignment Example



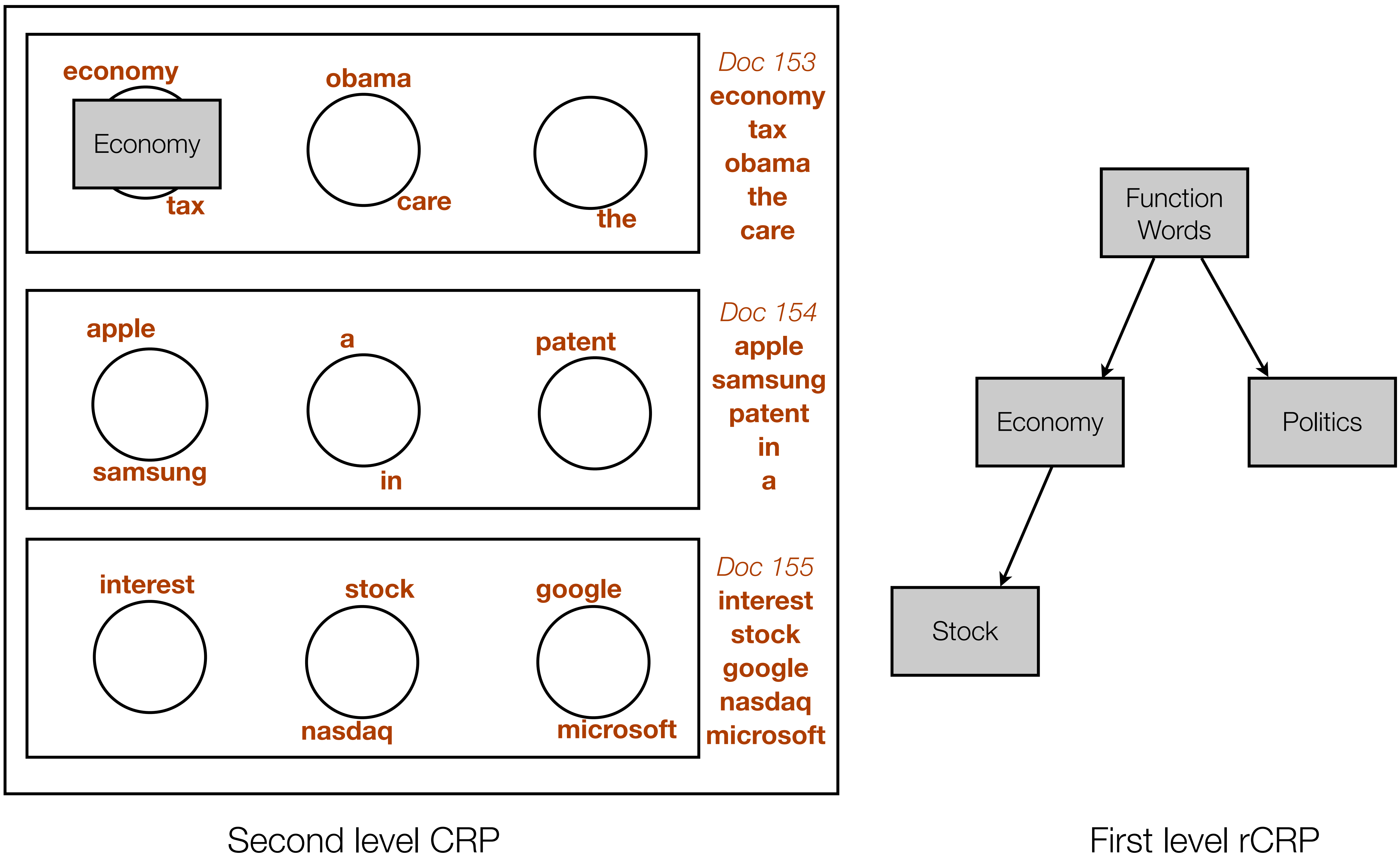


Our Model: Assignment Example

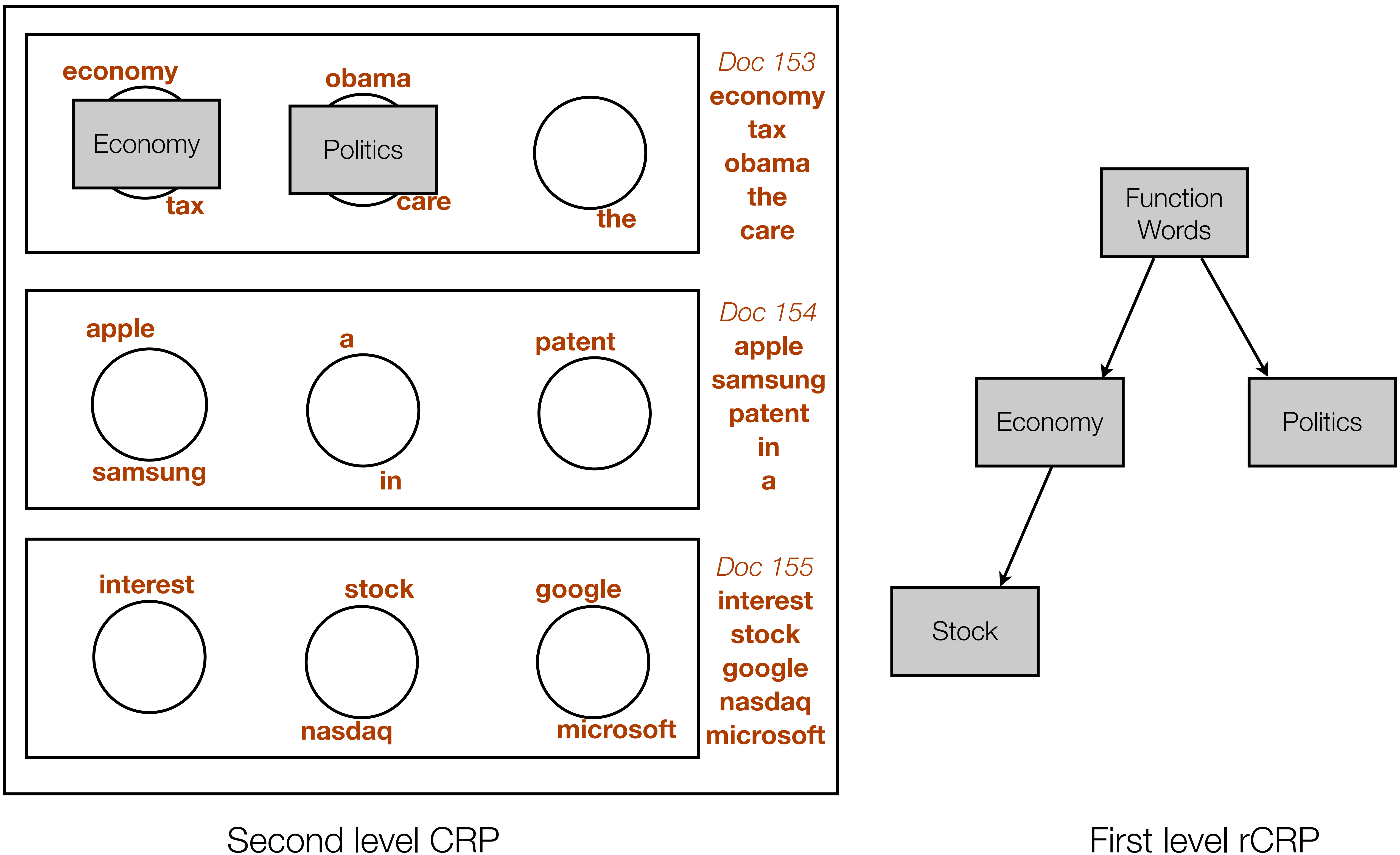




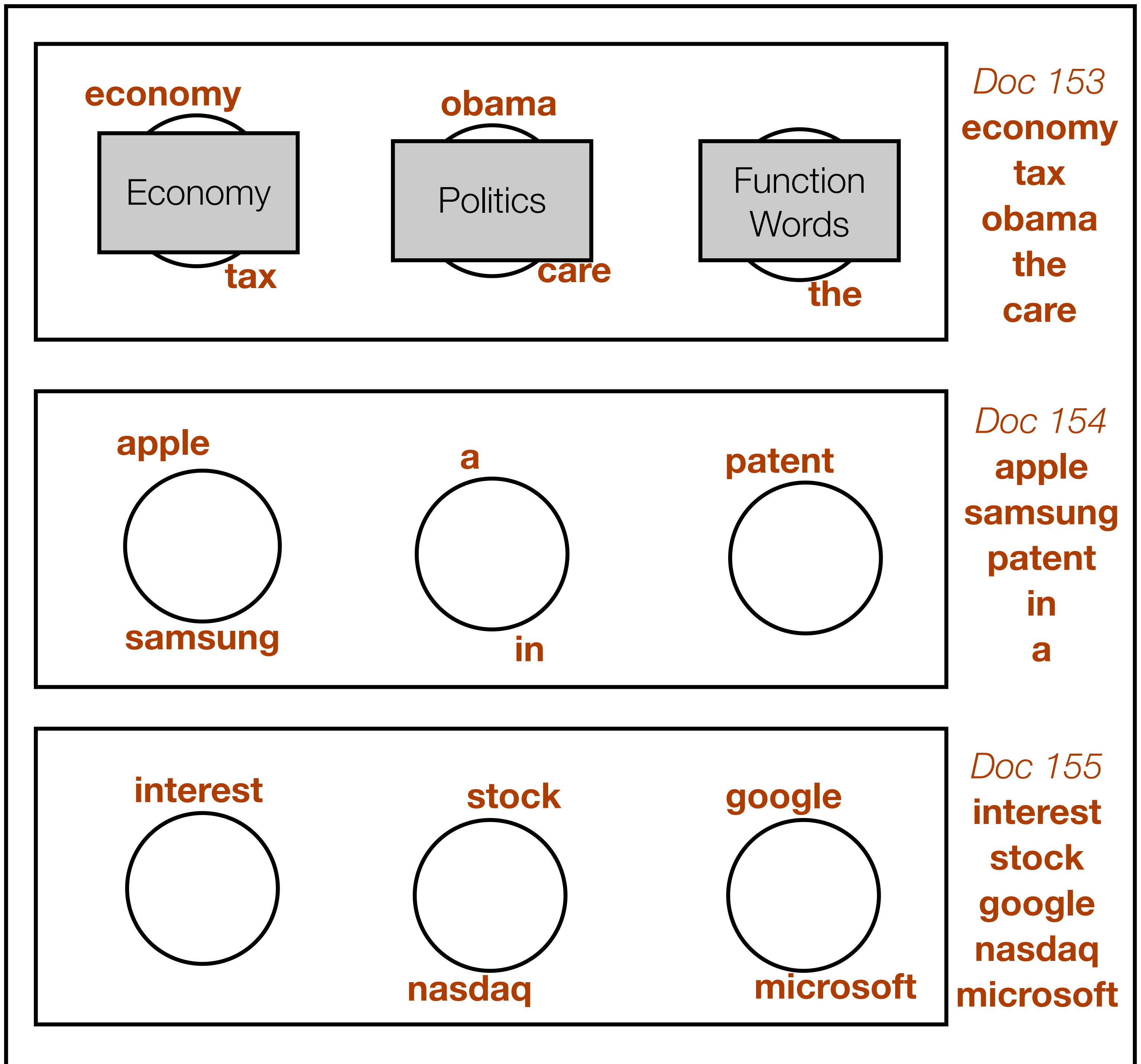
Our Model: Assignment Example



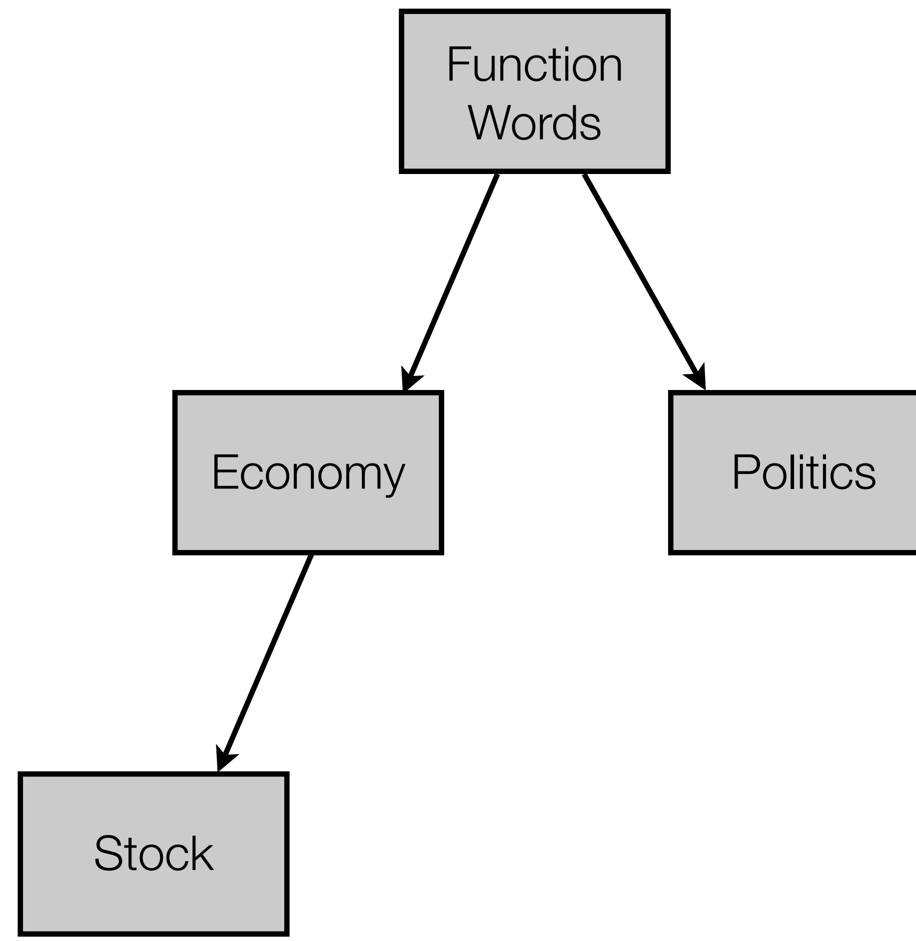
Our Model: Assignment Example



Our Model: Assignment Example

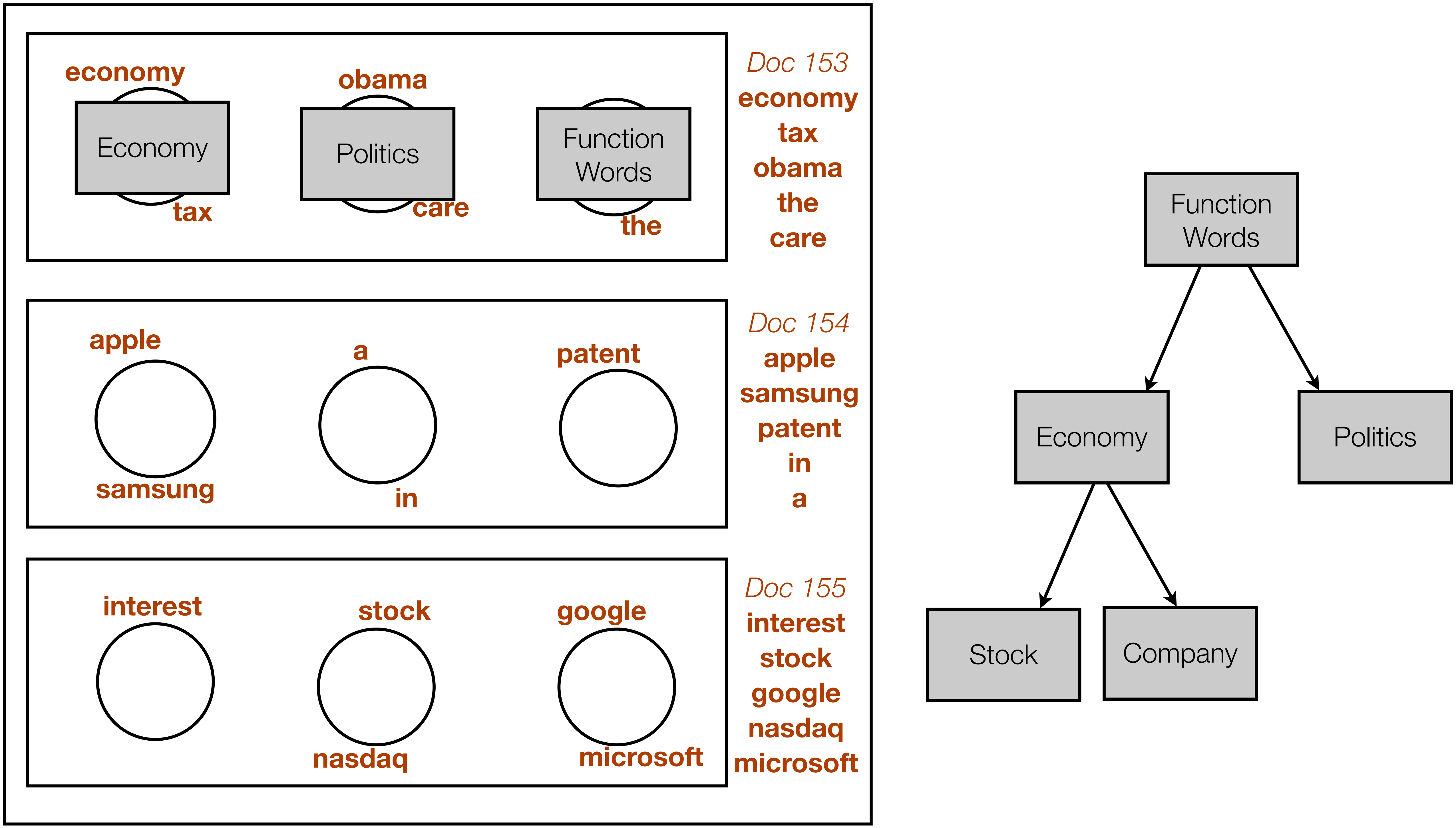


Second level CRP



First level rCRP

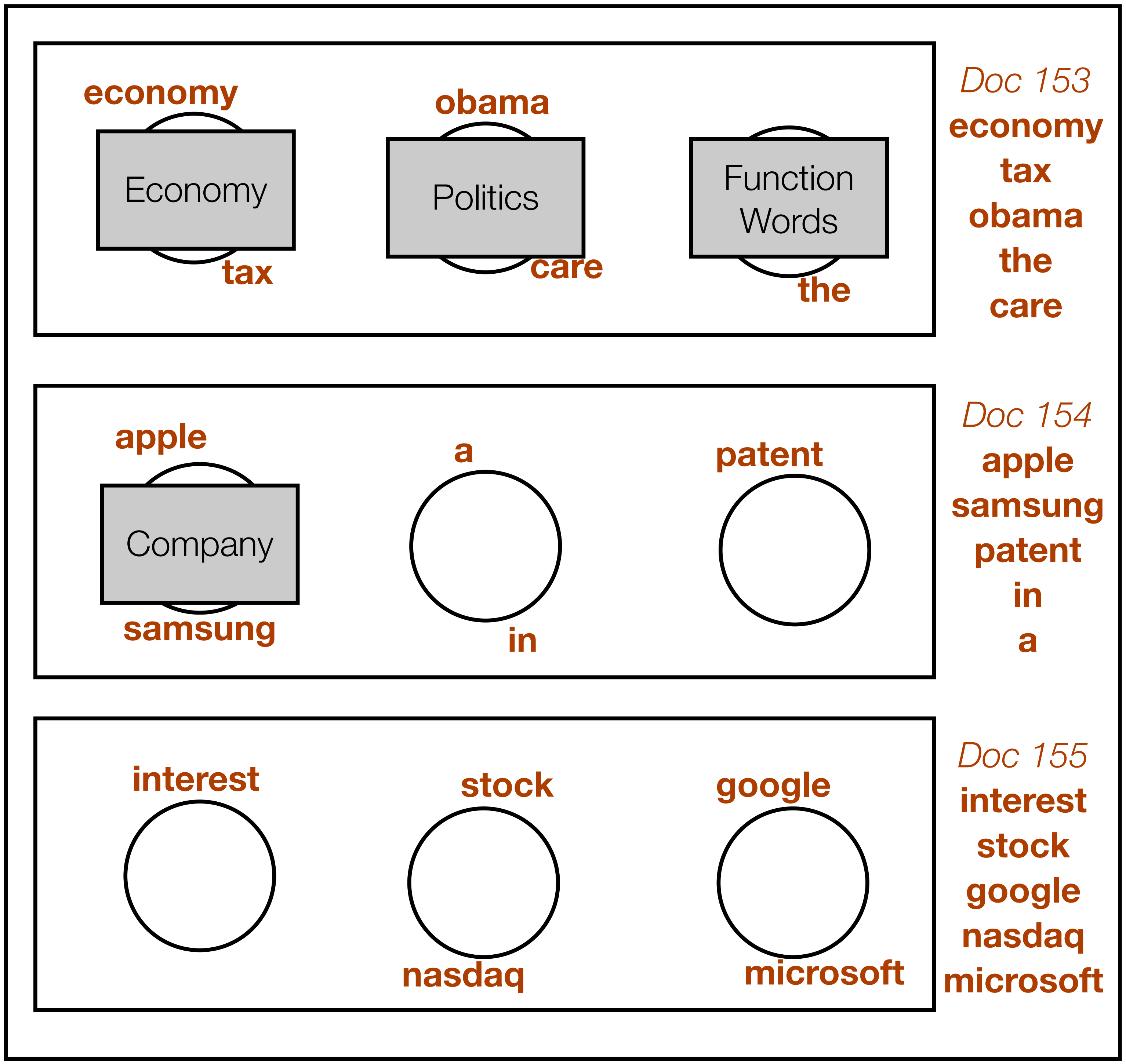
Our Model: Assignment Example



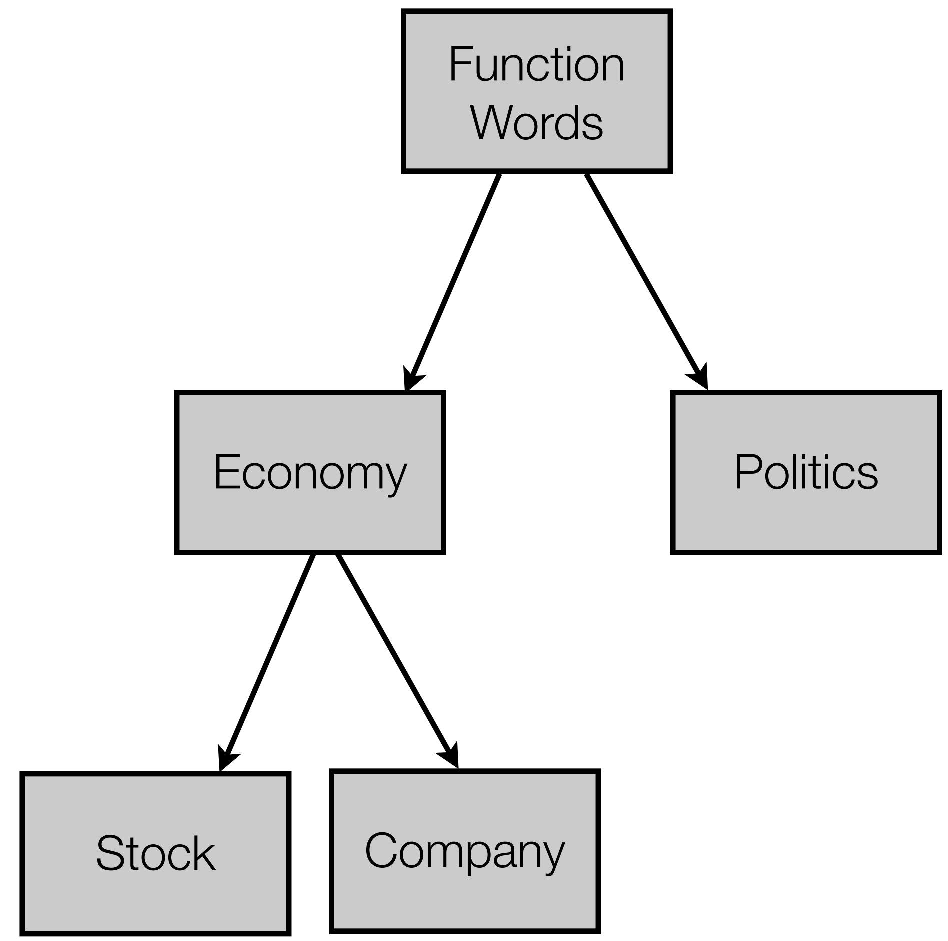
Second level CRP

First level rCRP

Our Model: Assignment Example

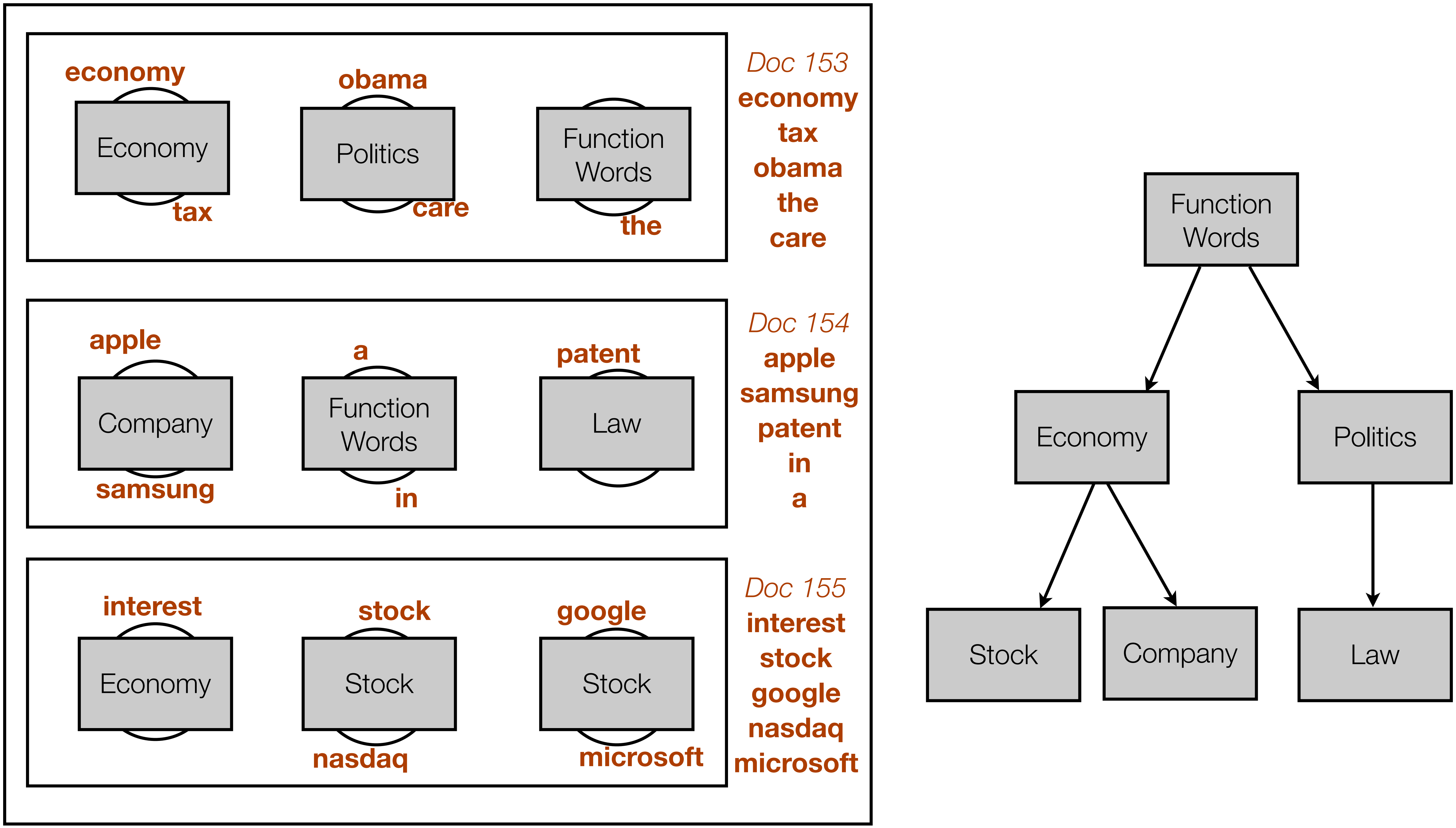


Second level CRP



First level rCRP

Our Model: Assignment Example

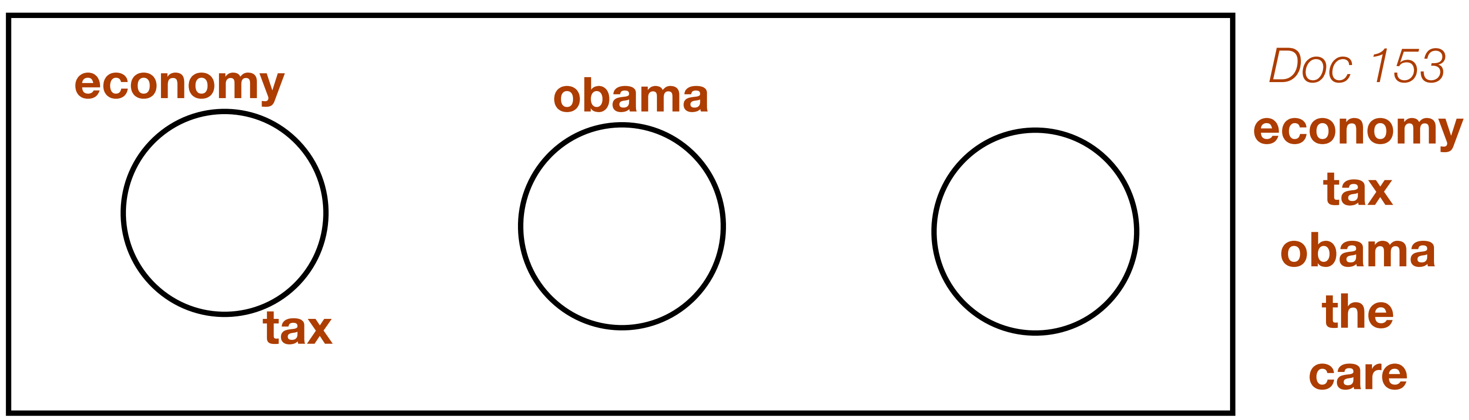


Second level CRP

First level rCRP

Our Model: Table Assignment Process

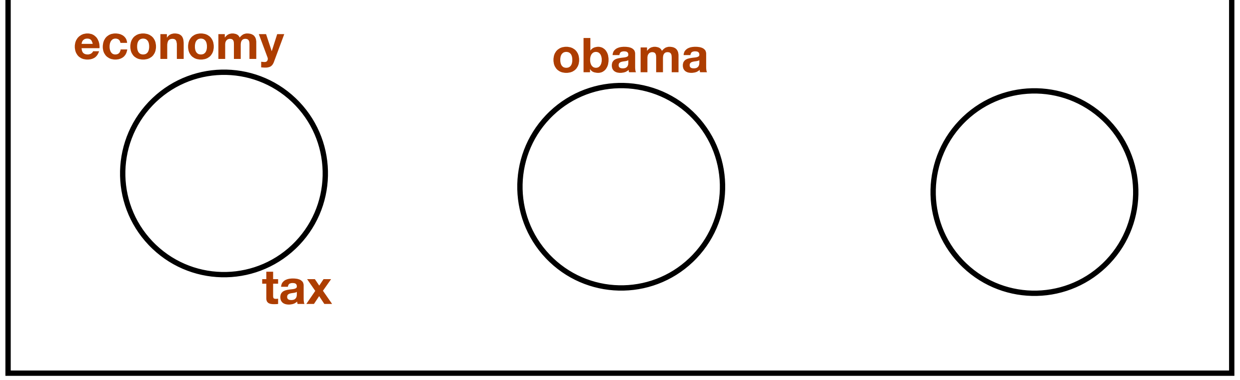
- Table assignment process
- A new customer sits at
  - an existing table $p \propto n_t$ where n_t = number of customers at table t
  - or a new table $p \propto \alpha$



Our Model: Table Assignment Process

- Table assignment process
- A new customer sits at
    - an existing table $p \propto n_t$ where n_t = number of customers at table t
    - or a new table $p \propto \alpha$

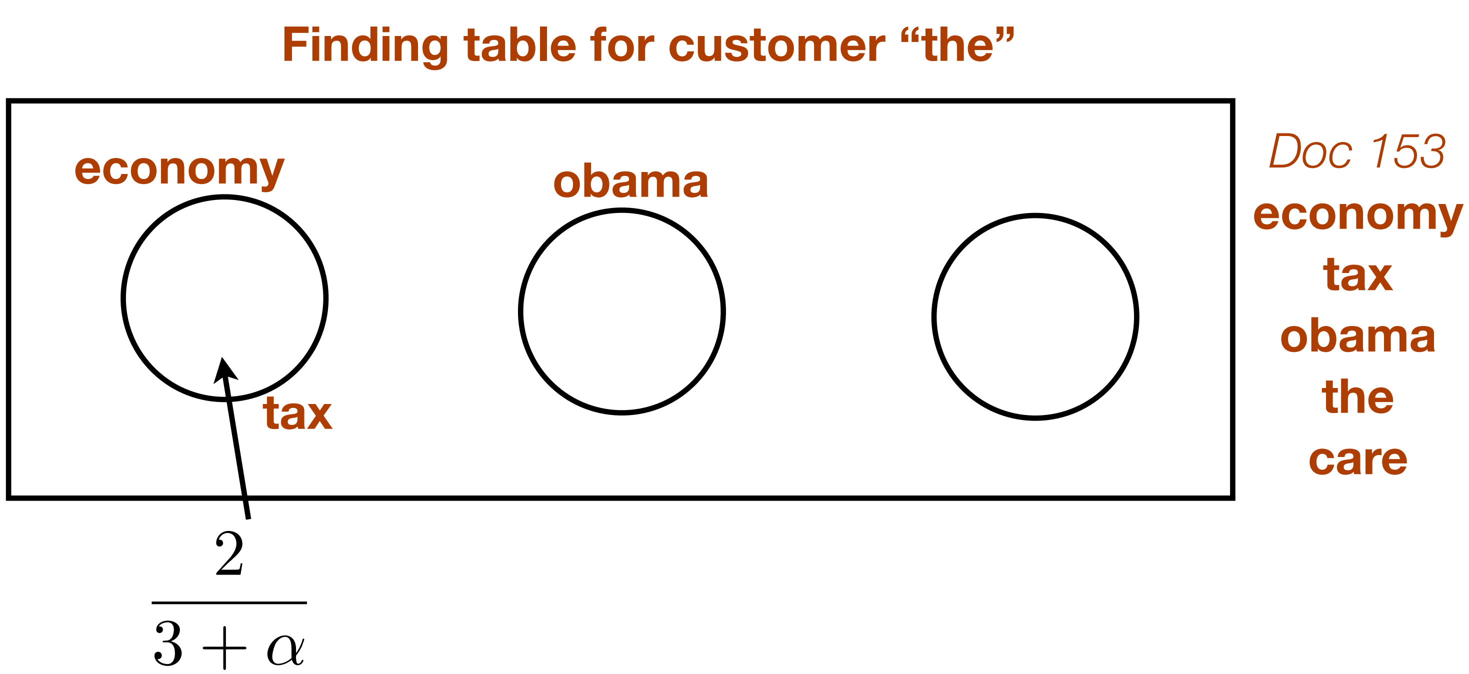
Finding table for customer "the"



Doc 153
economy
tax
obama
the
care

Our Model: Table Assignment Process

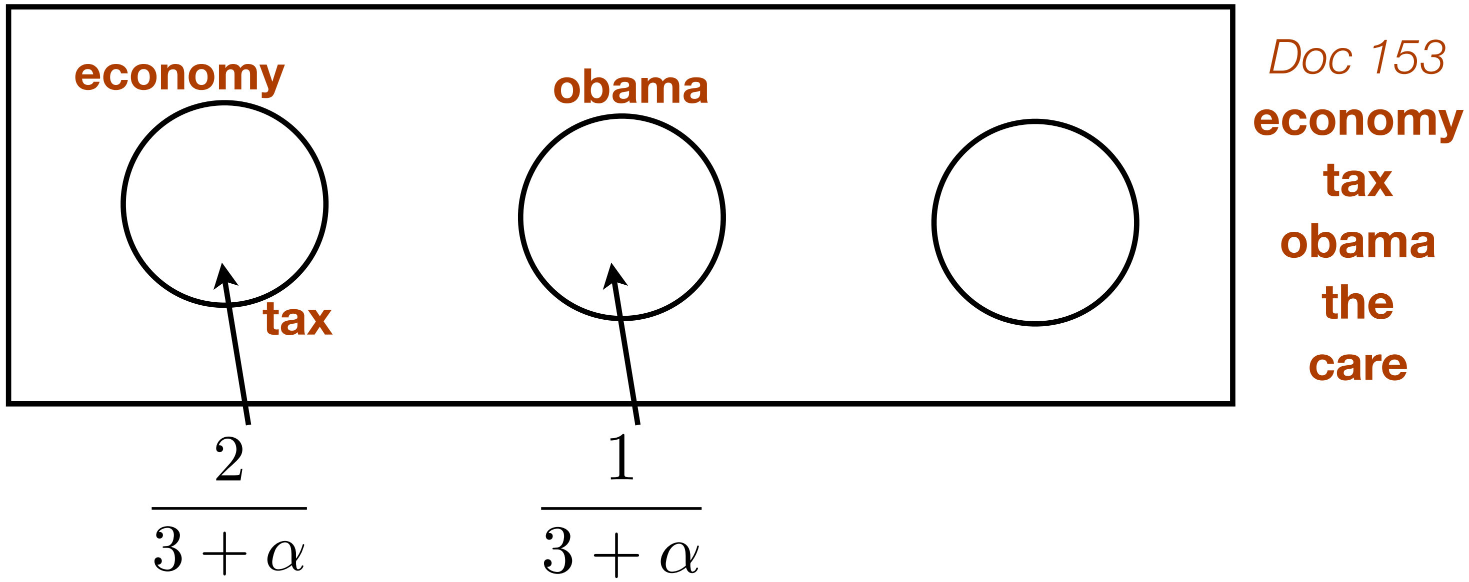
- Table assignment process
- A new customer sits at
  - an existing table $p \propto n_t$ where n_t = number of customers at table t
  - or a new table $p \propto \alpha$



Our Model: Table Assignment Process

- Table assignment process
- A new customer sits at
  - an existing table $p \propto n_t$ where n_t = number of customers at table t
  - or a new table $p \propto \alpha$

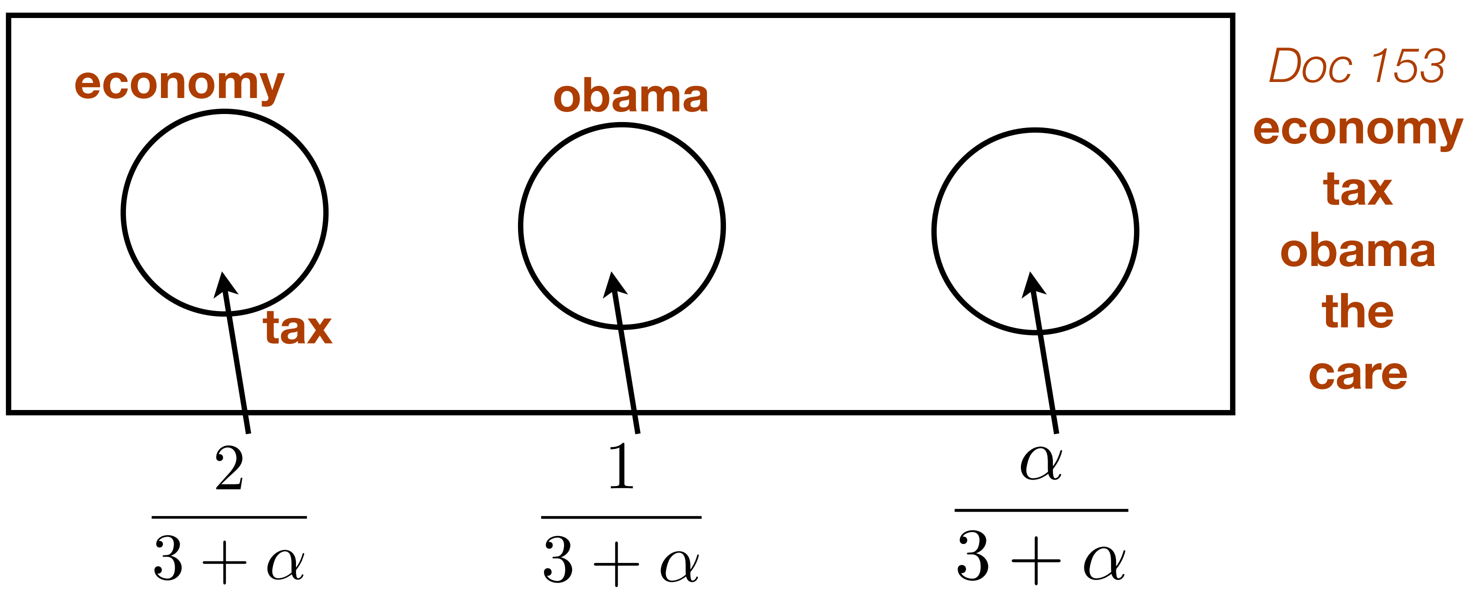
Finding table for customer "the"



Our Model: Table Assignment Process

- Table assignment process
- A new customer sits at
  - an existing table $p \propto n_t$ where n_t = number of customers at table t
  - or a new table $p \propto \alpha$

Finding table for customer "the"



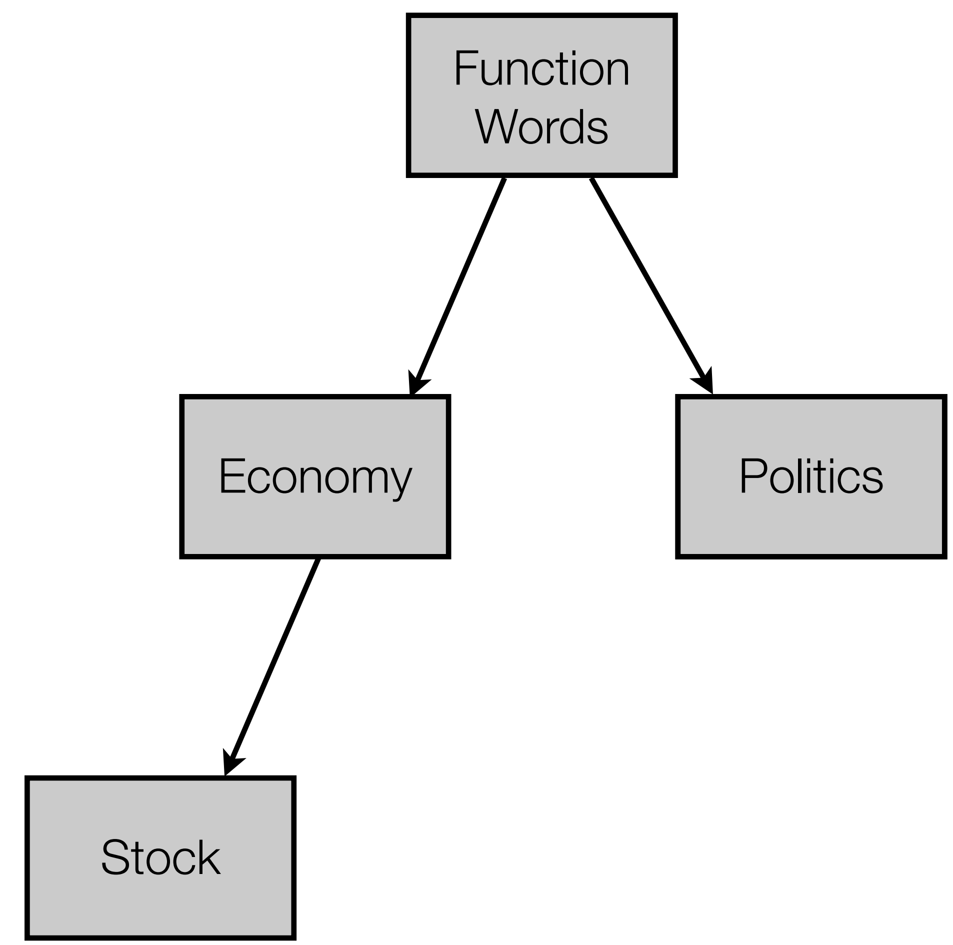
Our Model: Notation

- Dish assignment process
- rCPR, like CRP, show tendency of preferential attachment
- So it's important to keep count of tables serving a certain dish

- We use m to count the number of tables!
- m_{jk} = number of tables serving dish ϕ_k at restaurant j
- $m_{.k}$ = number of tables serving dish ϕ_k across all restaurants
- $m_{j.}$ = number of tables at restaurant j
- $M_{.k}$ = $m_{.k}$ summed over all its descents and itself

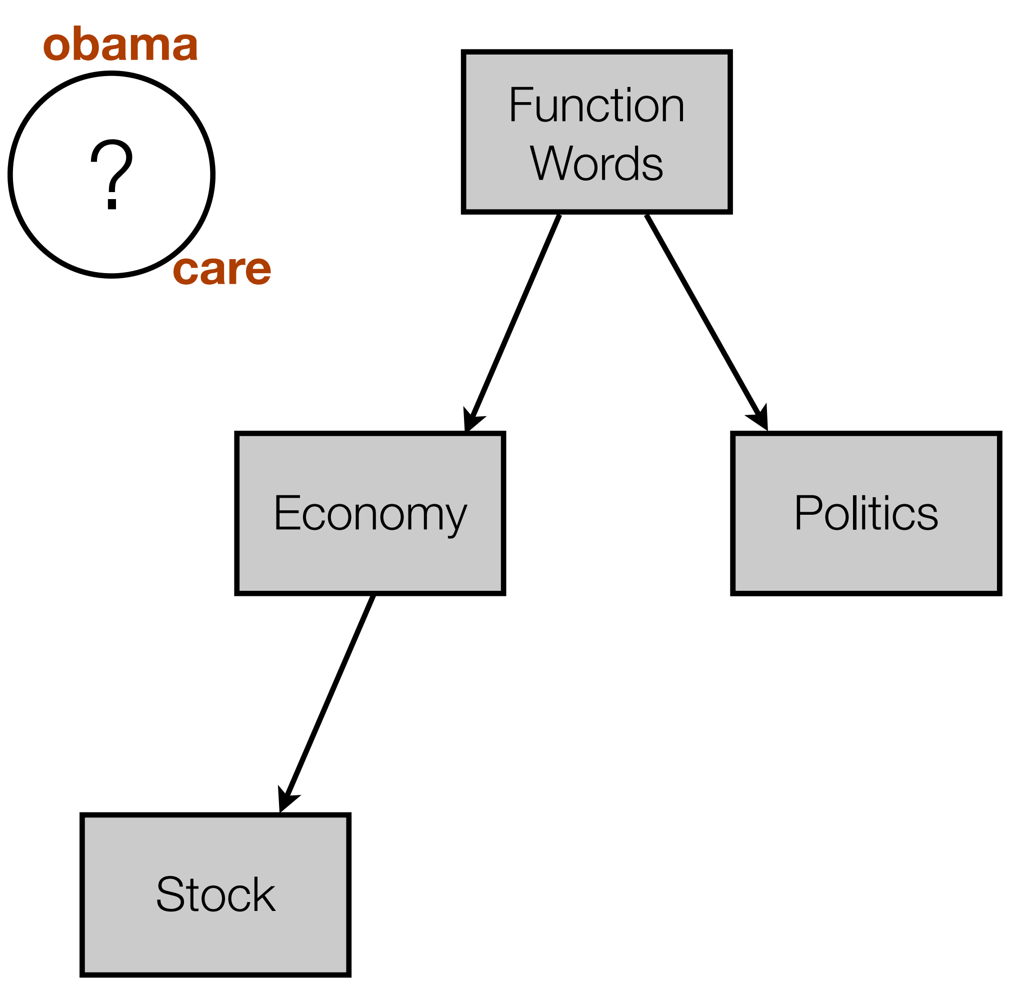
Our Model: Dish Assignment Process

- Dish assignment process by recursive search
- Start from the root dish, and move down the tree until we find the right dish
- Suppose that $\phi_{\boldsymbol{k}}$ is the current dish under examination
- We choose one of the three actions
  1. Stop at $\phi_{\boldsymbol{k}}$
     with probability proportion to $m_{.k}$
  2. Move down to one of the existing child dish $\phi_{\boldsymbol{k}'}$
     with probability proportion to $M_{.k'}$
  3. Move down and create a new child dish $\phi_{k_{\text{new}}}$
     with probability proportion to γ^n



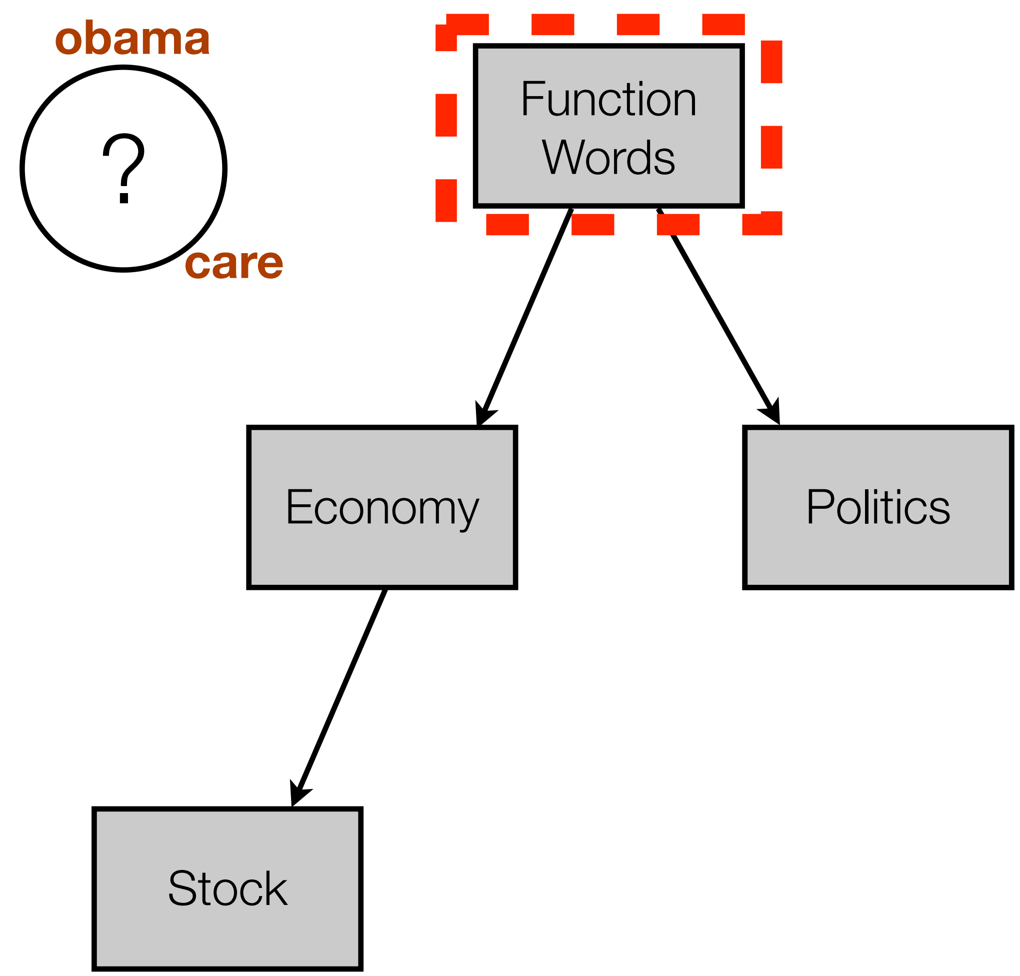
Our Model: Dish Assignment Process

- Dish assignment process by recursive search
- Start from the root dish, and move down the tree until we find the right dish
- Suppose that ϕ_k is the current dish under examination
- We choose one of the three actions
  1. Stop at ϕ_k
     with probability proportion to $m_{.k}$
  2. Move down to one of the existing child dish $\phi_{k'}$
     with probability proportion to $M_{.k'}$
  3. Move down and create a new child dish $\phi_{k_{\text{new}}}$
     with probability proportion to γ^n



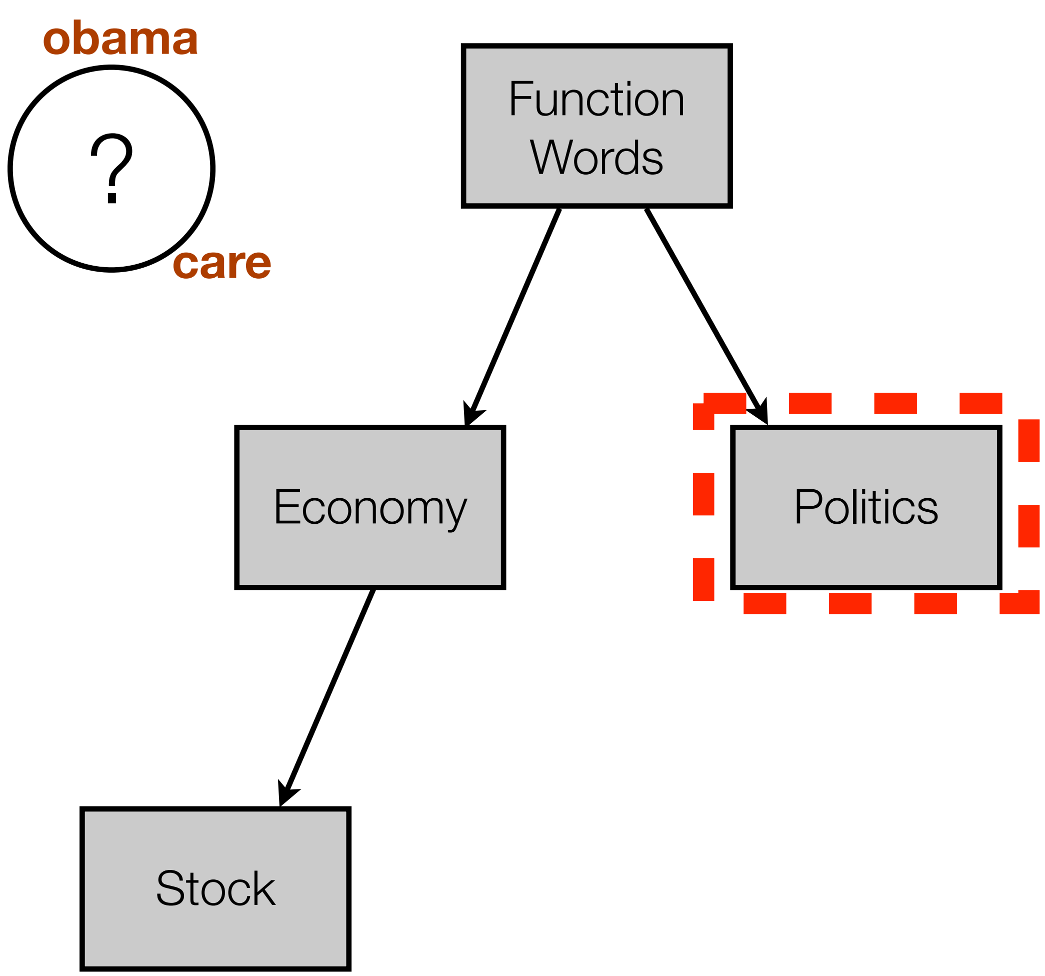
Our Model: Dish Assignment Process

- Dish assignment process by recursive search
- Start from the root dish, and move down the tree until we find the right dish
- Suppose that ϕ_k is the current dish under examination
- We choose one of the three actions
  1. Stop at ϕ_k
     with probability proportion to $m_{.k}$
  2. Move down to one of the existing child dish $\phi_{k'}$
     with probability proportion to $M_{.k'}$
  3. Move down and create a new child dish $\phi_{k_{\text{new}}}$
     with probability proportion to γ^n



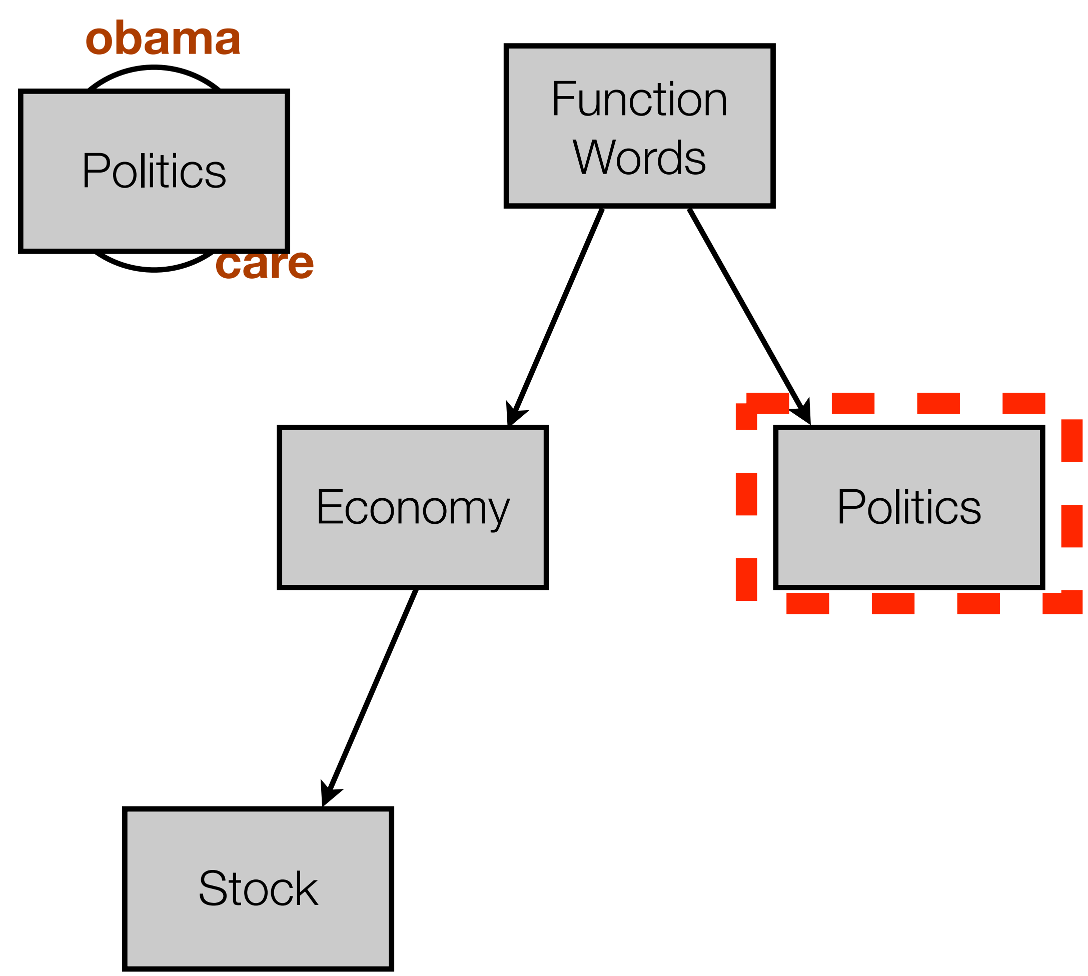
Our Model: Dish Assignment Process

- Dish assignment process by recursive search
- Start from the root dish, and move down the tree until we find the right dish
- Suppose that ϕ_k is the current dish under examination
- We choose one of the three actions
  1. Stop at ϕ_k
     with probability proportion to $m_{.k}$
  2. Move down to one of the existing child dish $\phi_{k'}$
     with probability proportion to $M_{.k'}$
  3. Move down and create a new child dish $\phi_{k_{\text{new}}}$
     with probability proportion to γ^n



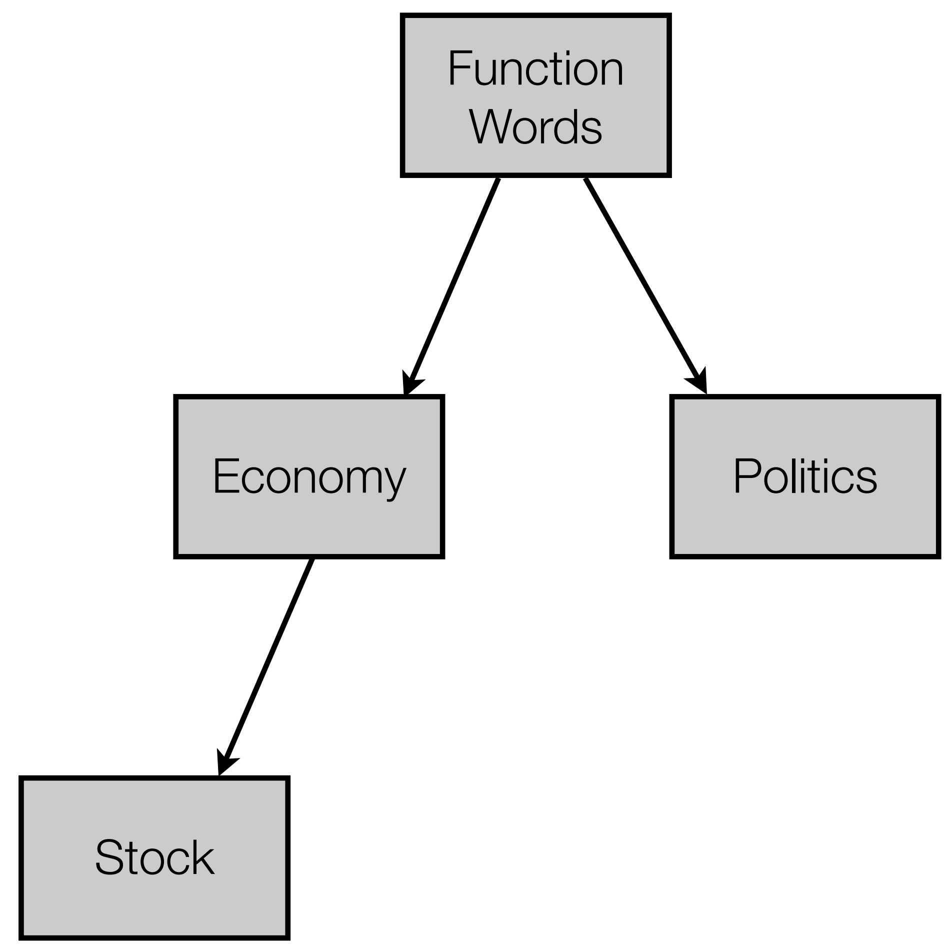
Our Model: Dish Assignment Process

- Dish assignment process by recursive search
- Start from the root dish, and move down the tree until we find the right dish
- Suppose that ϕ_k is the current dish under examination
- We choose one of the three actions
  1. Stop at ϕ_k with probability proportion to $m_{.k}$
  2. Move down to one of the existing child dish $\phi_{k'}$ with probability proportion to $M_{.k'}$
  3. Move down and create a new child dish $\phi_{k_{\text{new}}}$ with probability proportion to γ^n



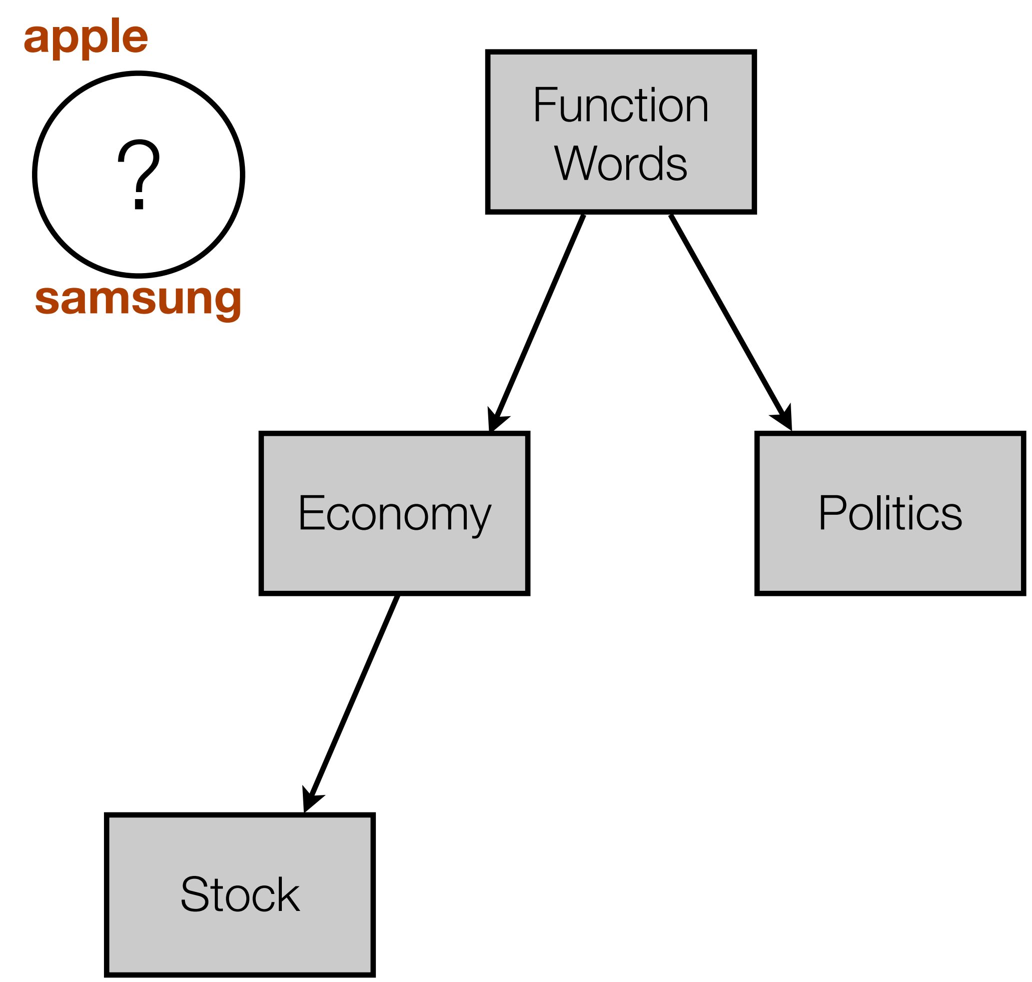
Our Model: Dish Assignment Process

- Dish assignment process by recursive search
- Start from the root dish, and move down the tree until we find the right dish
- Suppose that $\phi_{\boldsymbol{k}}$ is the current dish under examination
- We choose one of the three actions
  1. Stop at $\phi_{\boldsymbol{k}}$
     with probability proportion to $m_{.k}$
  2. Move down to one of the existing child dish $\phi_{\boldsymbol{k}'}$
     with probability proportion to $M_{.k'}$
  3. Move down and create a new child dish $\phi_{\boldsymbol{k}_{\text{new}}}$
     with probability proportion to γ^n



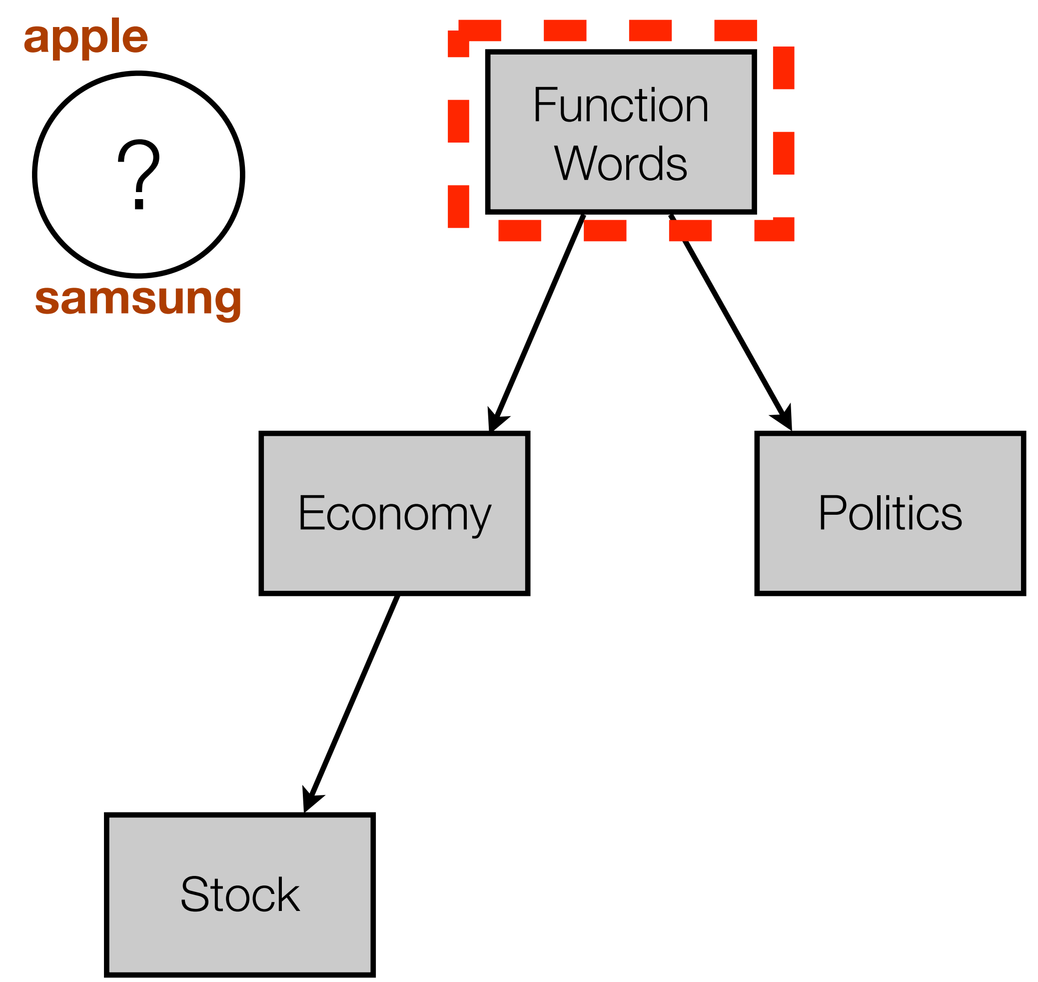
Our Model: Dish Assignment Process

- Dish assignment process by recursive search
- Start from the root dish, and move down the tree until we find the right dish
- Suppose that ϕ_k is the current dish under examination
- We choose one of the three actions
  1. Stop at ϕ_k
     with probability proportion to $m_{.k}$
  2. Move down to one of the existing child dish $\phi_{k'}$
     with probability proportion to $M_{.k'}$
  3. Move down and create a new child dish $\phi_{k_{\mathrm{new}}}$
     with probability proportion to γ^n



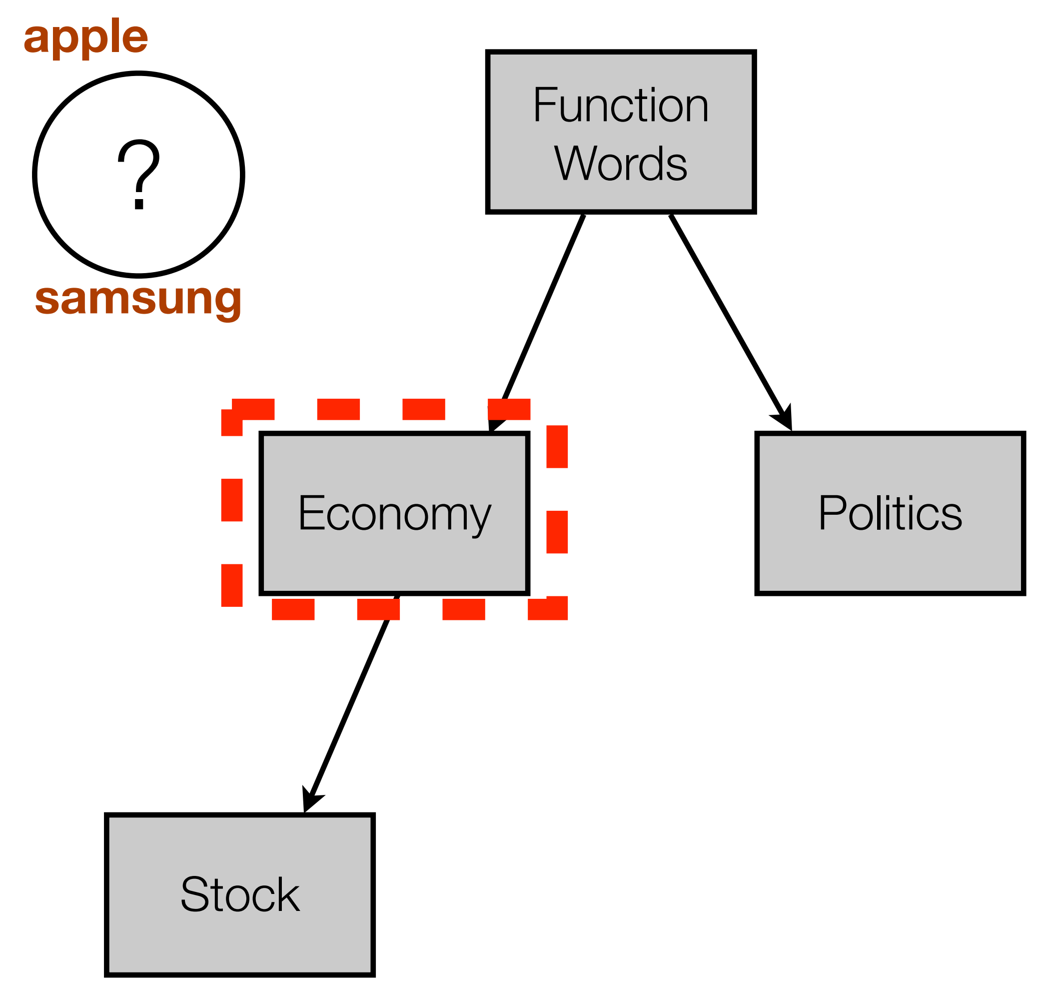
Our Model: Dish Assignment Process

- Dish assignment process by recursive search
- Start from the root dish, and move down the tree until we find the right dish
- Suppose that ϕ_k is the current dish under examination
- We choose one of the three actions
  1. Stop at ϕ_k with probability proportion to $m_{.k}$
  2. Move down to one of the existing child dish $\phi_{k'}$ with probability proportion to $M_{.k'}$
  3. Move down and create a new child dish $\phi_{k_{\text{new}}}$ with probability proportion to γ^n



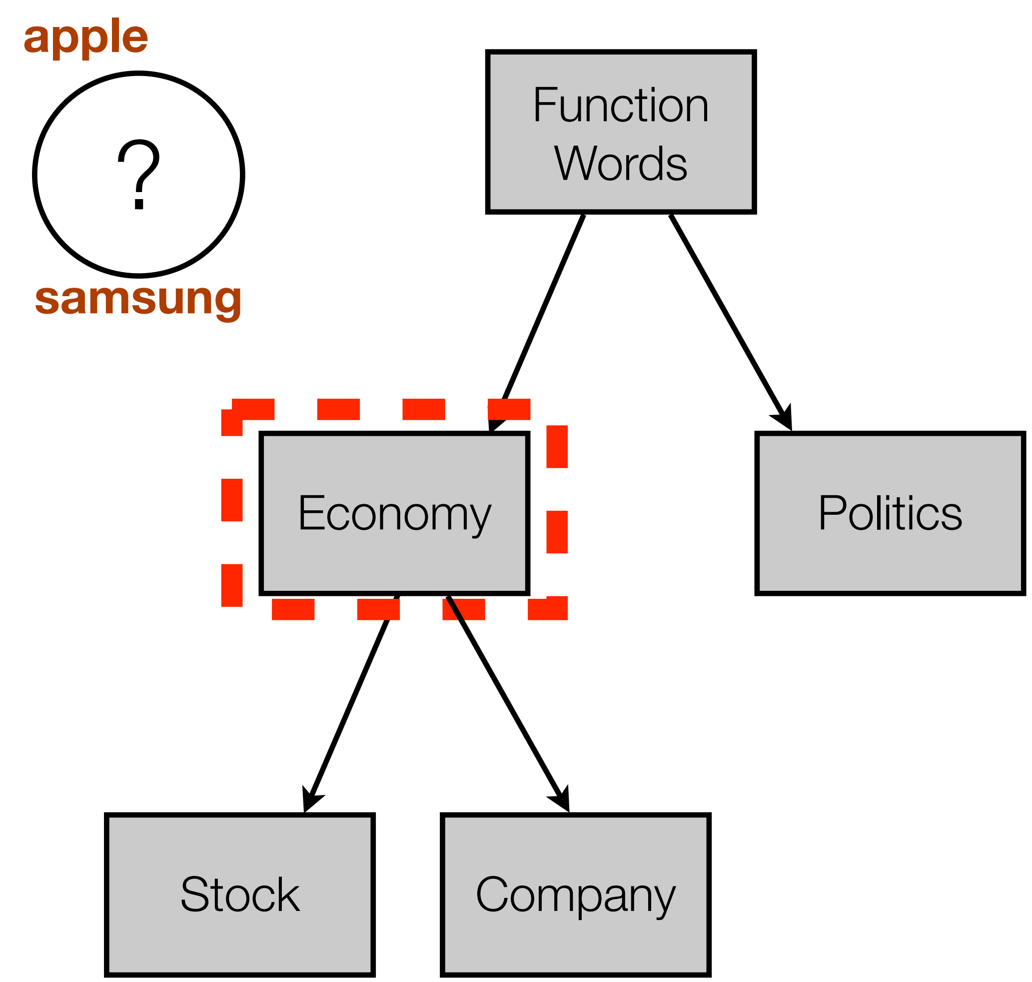
Our Model: Dish Assignment Process

- Dish assignment process by recursive search
- Start from the root dish, and move down the tree until we find the right dish
- Suppose that ϕ_k is the current dish under examination
- We choose one of the three actions
  1. Stop at ϕ_k
     with probability proportion to $m_{.k}$
  2. Move down to one of the existing child dish $\phi_{k'}$
     with probability proportion to $M_{.k'}$
  3. Move down and create a new child dish $\phi_{k_{\mathrm{new}}}$
     with probability proportion to γ^n



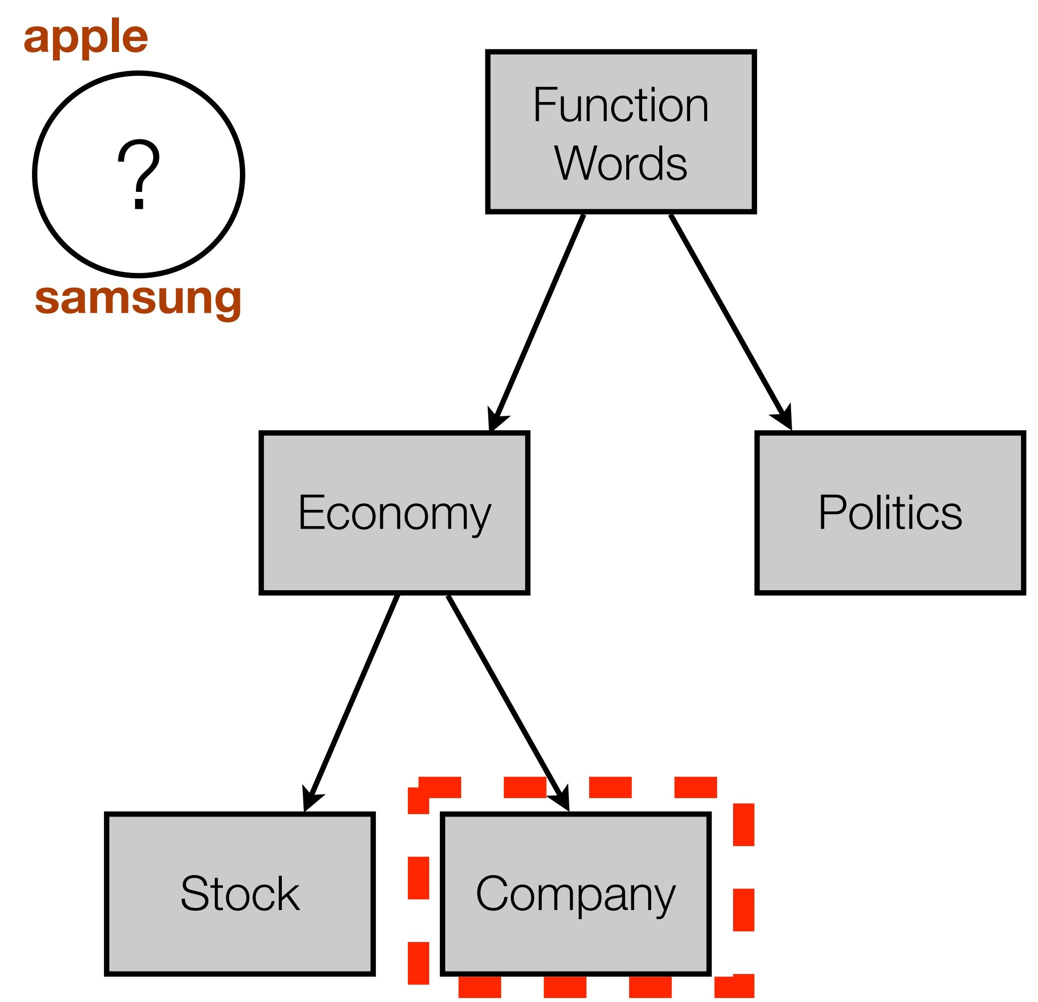
Our Model: Dish Assignment Process

- Dish assignment process by recursive search
- Start from the root dish, and move down the tree until we find the right dish
- Suppose that ϕ_k is the current dish under examination
- We choose one of the three actions
    1. Stop at ϕ_k
       with probability proportion to $m_{.k}$
    2. Move down to one of the existing child dish $\phi_{k'}$
       with probability proportion to $M_{.k'}$
    3. Move down and create a new child dish $\phi_{k_{\text{new}}}$
       with probability proportion to γ^n



Our Model: Dish Assignment Process

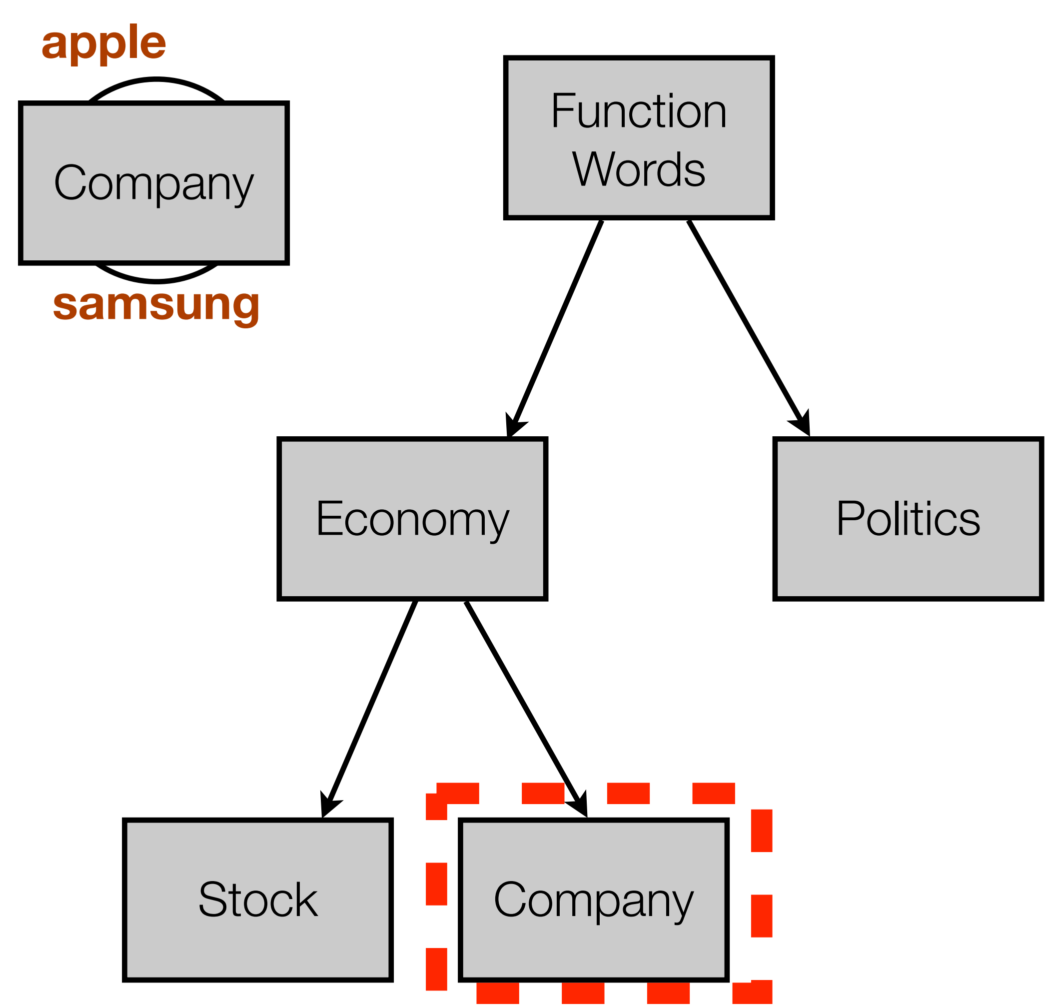
- Dish assignment process by recursive search
- Start from the root dish, and move down the tree until we find the right dish
- Suppose that ϕ_k is the current dish under examination
- We choose one of the three actions
  1. Stop at ϕ_k
     with probability proportion to $m_{.k}$
  2. Move down to one of the existing child dish $\phi_{k'}$
     with probability proportion to $M_{.k'}$
  3. Move down and create a new child dish $\phi_{k_{\text{new}}}$
     with probability proportion to γ^n



Our Model: Dish Assignment Process

- Dish assignment process by recursive search
- Start from the root dish, and move down the tree until we find the right dish
- Suppose that ϕ_k is the current dish under examination
- We choose one of the three actions

  1. Stop at ϕ_k with probability proportion to $m_{\cdot k}$
  2. Move down to one of the existing child dish $\phi_{k'}$ with probability proportion to $M_{\cdot k'}$
  3. Move down and create a new child dish $\phi_{k_{\mathrm{new}}}$ with probability proportion to γ^n



Posterior Inference

- Gibbs sampling with Markov Chain Monte Carlo
- Sampling table assignment

$$p(t_{ji} = t|rest) \propto \begin{cases} n_{jt}^{-ji} \times p(x_{ji}|\mathbf{x}^{-ji}, \mathbf{t}^{-ji}, t_{ji} = t, \mathbf{k}) \\ \alpha \times p(x_{ji}|\mathbf{x}^{-ji}, \mathbf{t}^{-ji}, t_{ji} = t_{\text{new}}, \mathbf{k}) \end{cases}$$

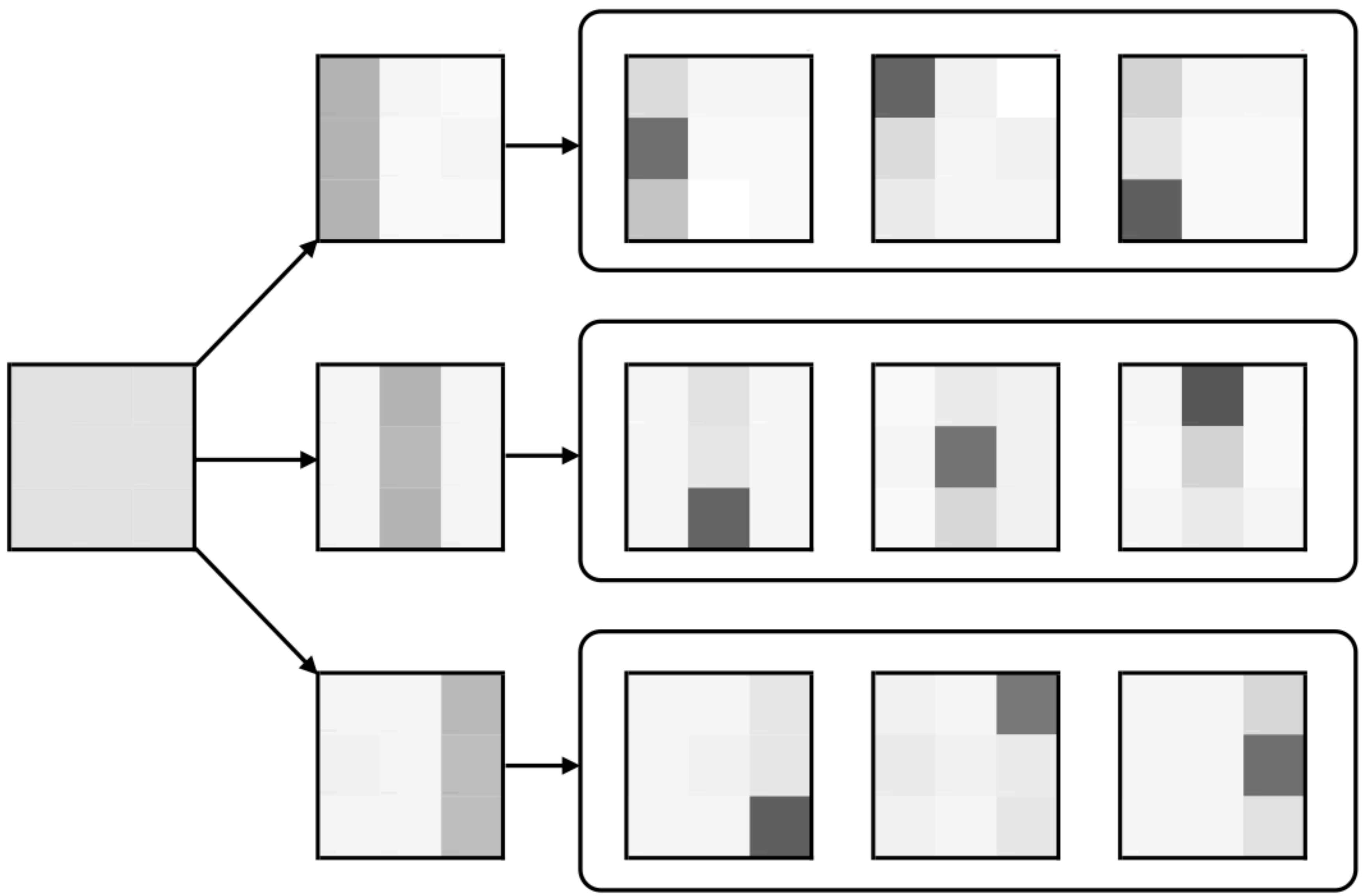
- Sampling dish assignment

$$p(k_{jt} = k|t, k^{-jt}, k_{\text{current}})$$
$$\propto \begin{cases} m_{.k}^{-jt} \times p(\mathbf{x}_{jt}|\mathbf{x}^{-jt}, \mathbf{t}, \mathbf{k}) \text{ if } k = k_{\text{current}} \\ M_{.k}^{-jt} \times p(\mathbf{x}_{jt}|\mathbf{x}^{-jt}, \mathbf{t}, \mathbf{k}) \text{ if } k = \text{a child of } k_{\text{current}} \\ \gamma^{n} \times p(\mathbf{x}_{jt}|\mathbf{x}^{-jt}, \mathbf{t}, \mathbf{k}) \text{ if } k = \text{a new child of } k_{\text{current}} \end{cases}$$

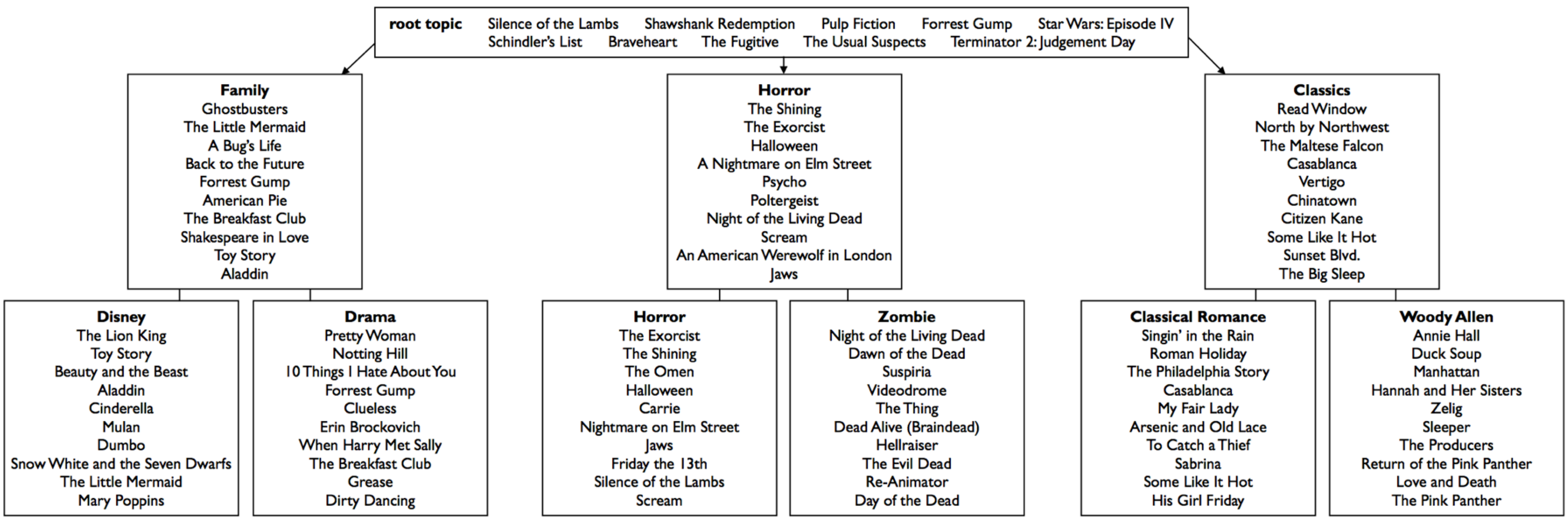
Experiments: Data Statistics

	# Documents	# Unique Word	Avg. Doc Length
Synthetic Data	1000	9	1000
New York Times	10000	6841	1886
Movie Lens	71567	10681	56
Wikipedia Contemporary Art	3600	6386	445

Experiments: Synthetic Data Test

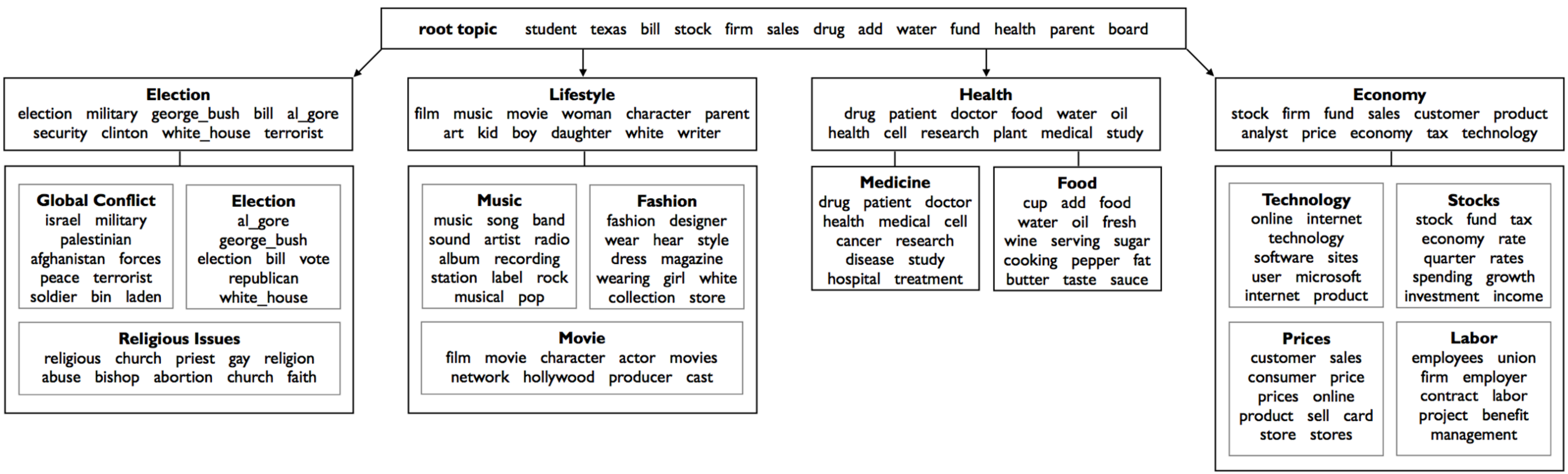


Experiments: Qualitative Result



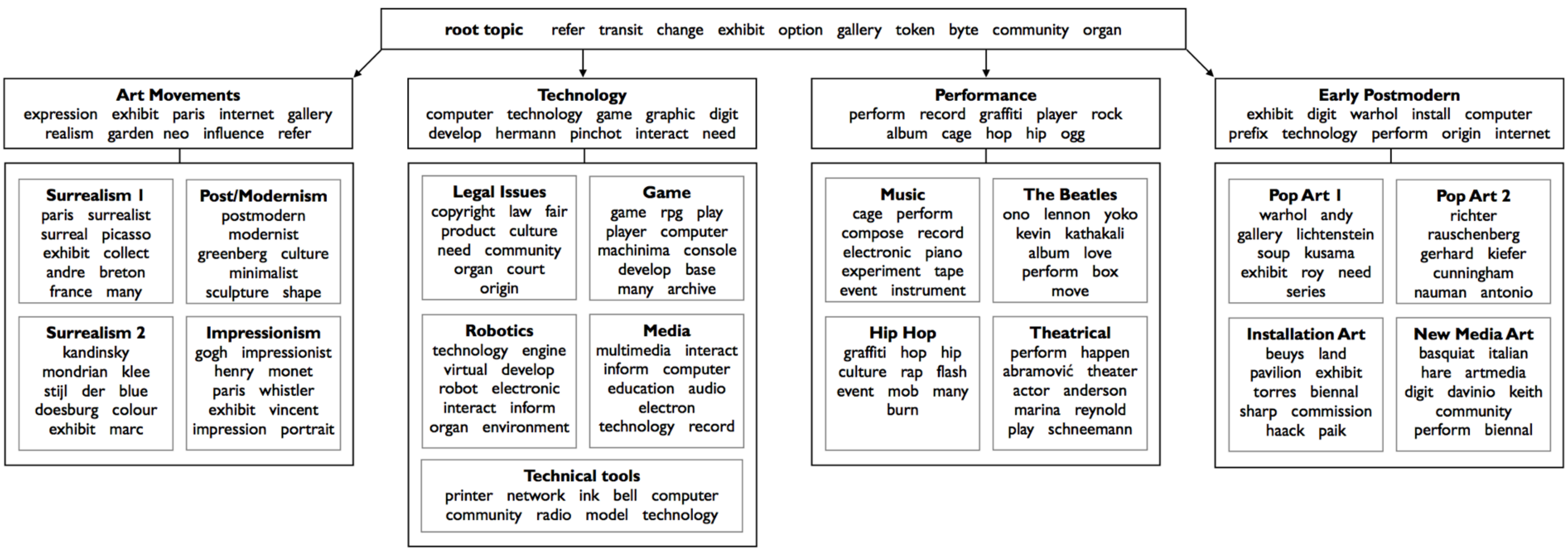
(a) MovieLens

Experiments: Qualitative Result



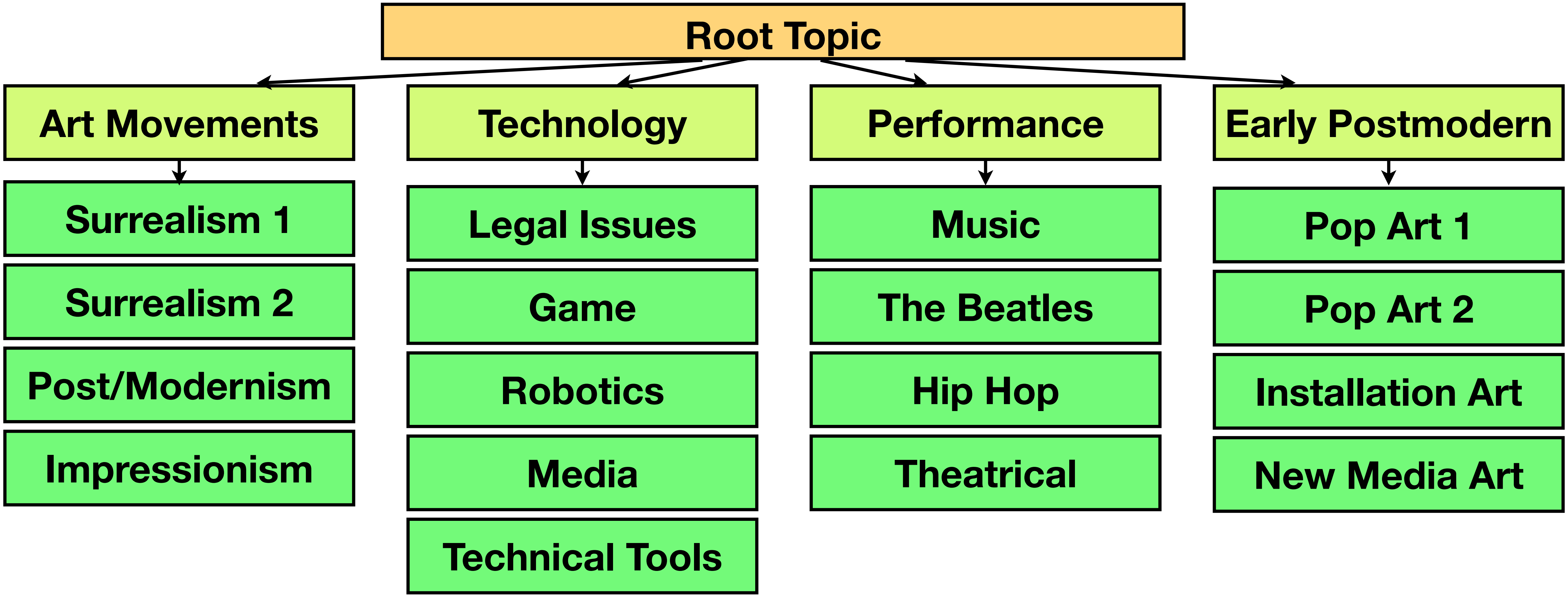
(b) New York Times

Experiments: Qualitative Result

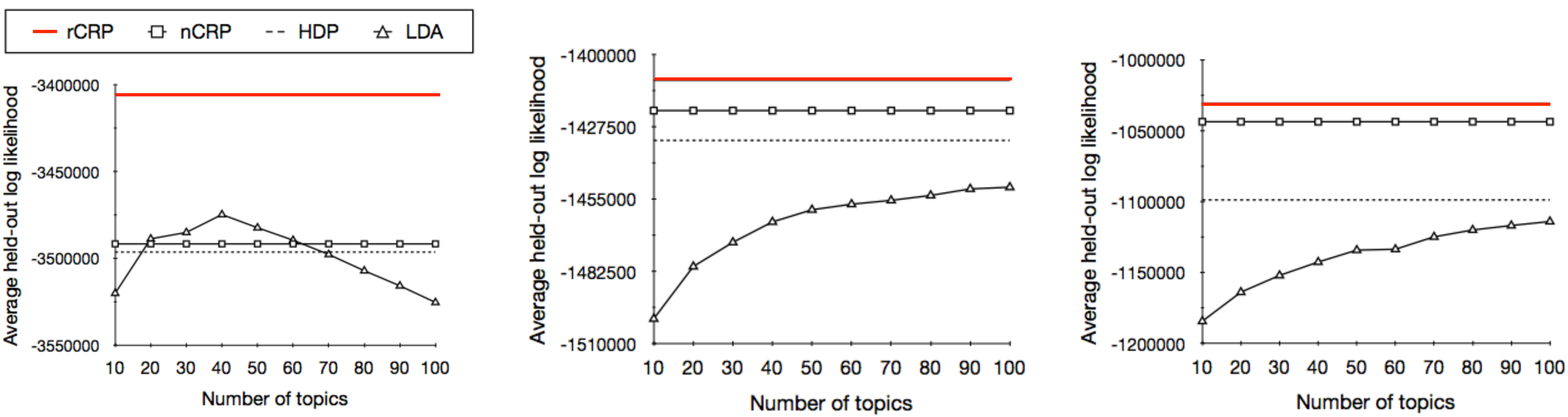


(c) Wikipedia

Experiments: Qualitative Result



Experiments: Held-out Likelihood

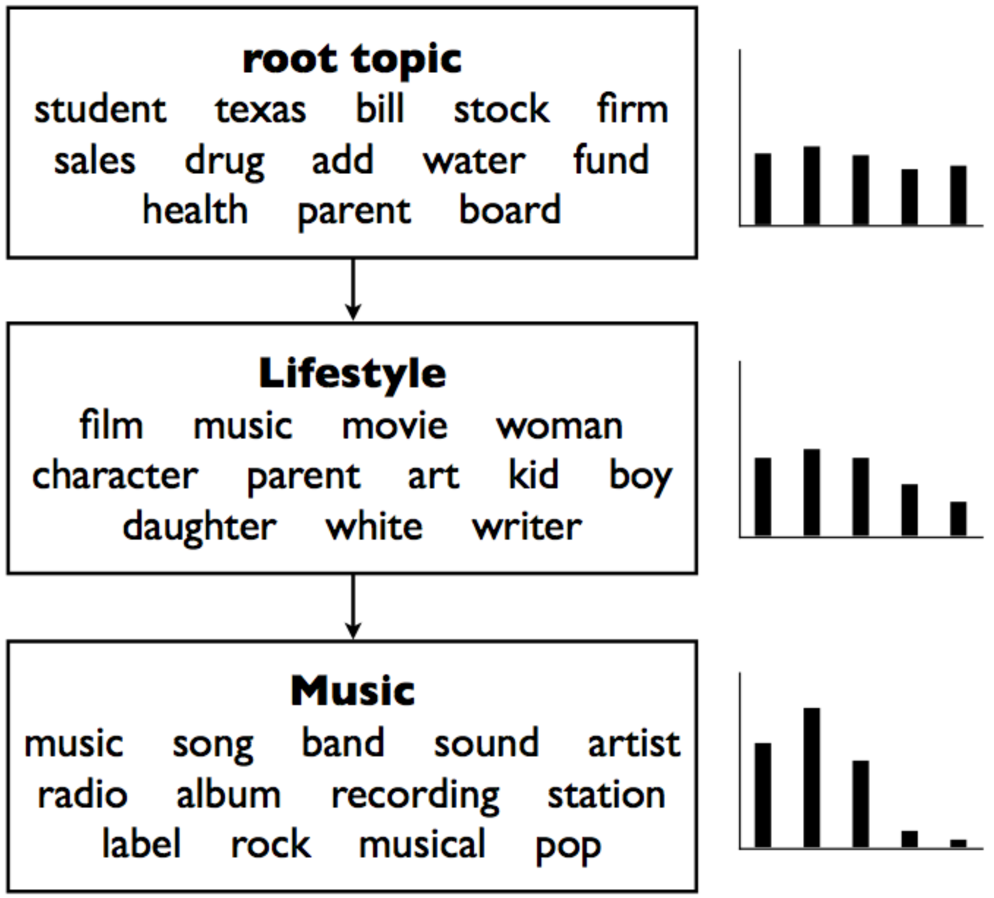


(a) MovieLens (b) New York Times (c) Wikipedia

Held-out likelihood
rCRP performs better than all other models - nCRP, HDP, LDA - in terms of the held-out likelihood

Experiments: Hierarchy Analysis 1

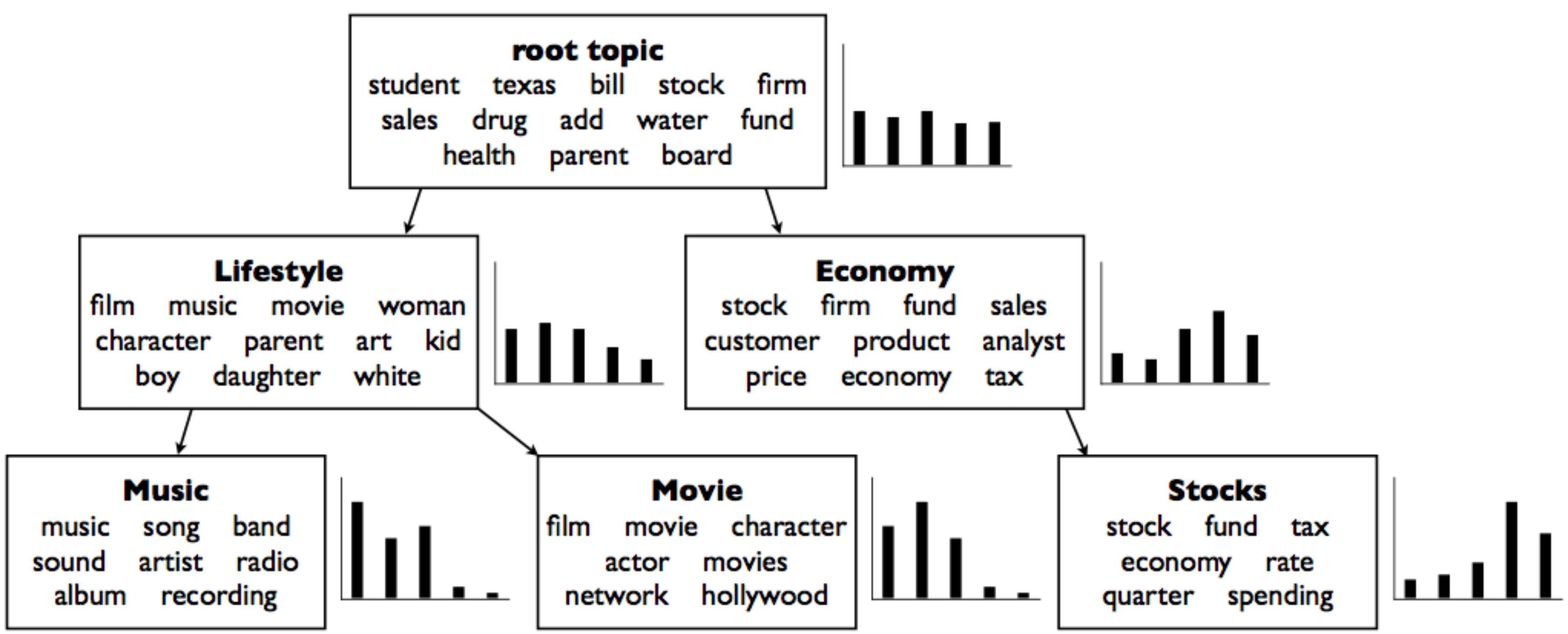
- **Topic Specialization**
  - We want to discover both general topics and specific topics
  - Ideally, root topic is the most general, and topics become more specific as we move down the topic tree



- Measure how much a topic has become specialized by **cosine distance from the norm**

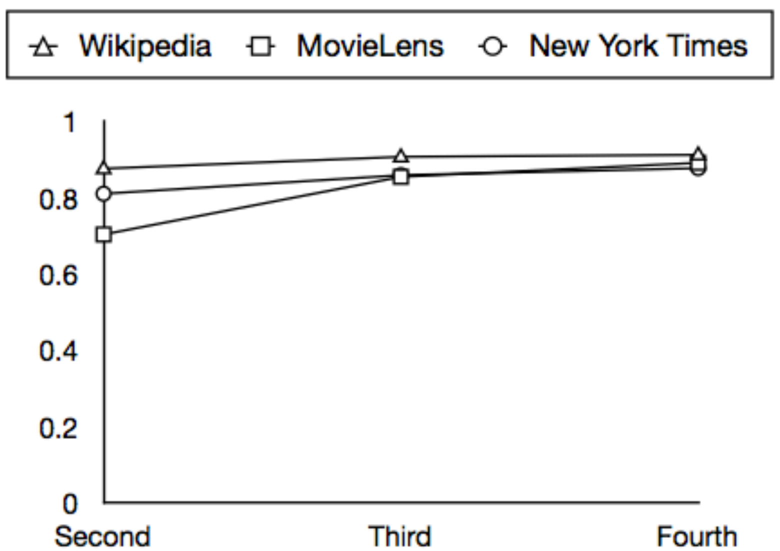
Experiments: Hierarchy Analysis 1

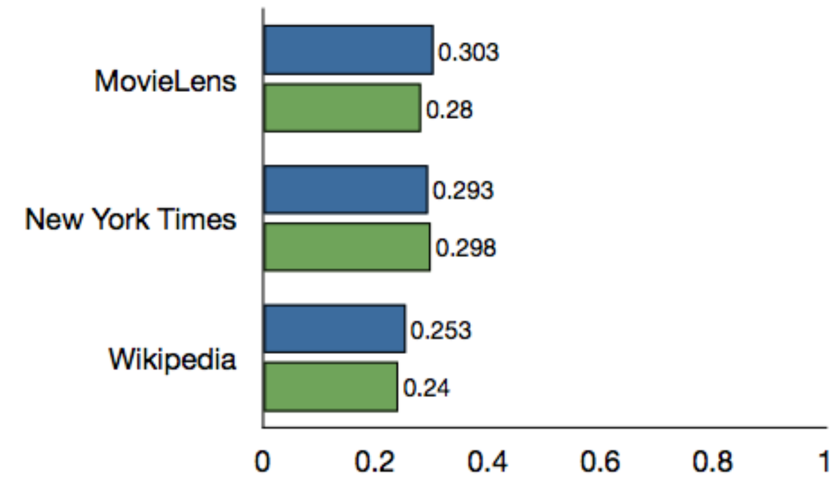
- **Hierarchical Affinity**
  - Topic pairs that form parent-child relation should be more similar
  - For clarity, limit to second (parent) and third (child) level topics
  - Compare **average cosine similarity** between topic pairs that form parent-child relation and those that don’t



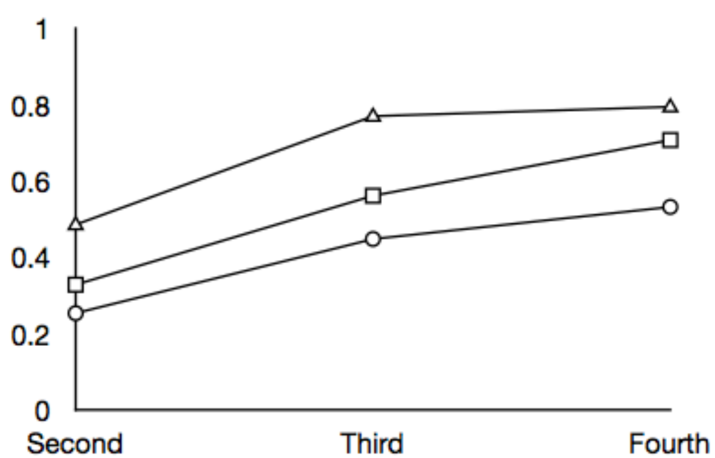
Experiments: Hierarchy Analysis Result

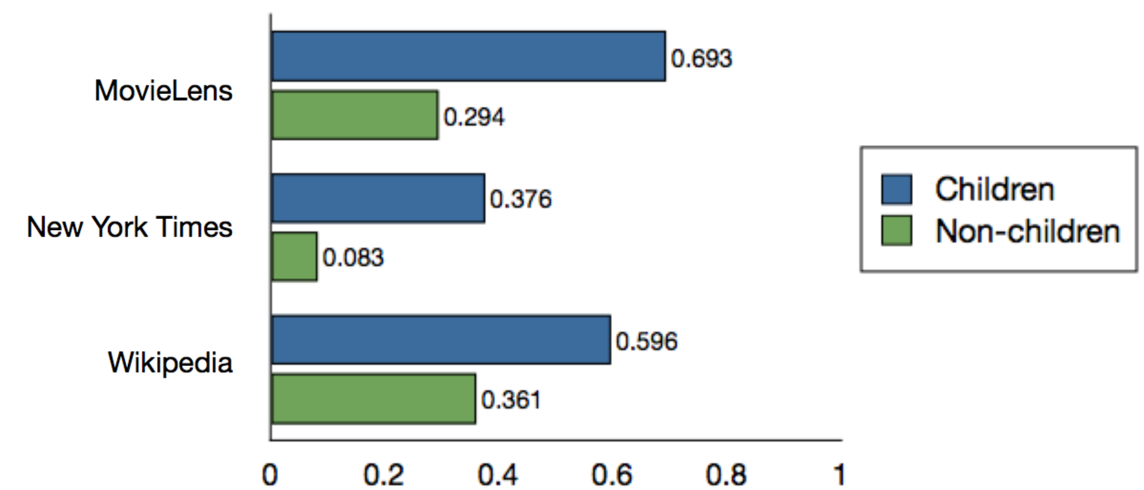
nCRP





rCRP





Topic Specialization
In rCRP as we move down the topic tree, topics become more specialized. In nCRP such tendency is not significant.

Hierarchical Affinity
In rCRP topics that form parent-child relation are more similar than those that don't. In nCRP such tendency is not significant.

Contribution

- Propose new nonparametric prior, rCRP
  - The model learns topic hierarchy from unstructured documents
  - Topic distribution of document is unlimited
  - Structure of topic tree is very flexible
- Define two new metrics for measuring the goodness of topic hierarchy
  - Topic specialization
  - Hierarchical affinity

Questions?