

# Refined Experts

Improving Classification in Large Hierarchies

*SIGIR 2009*

July 20, 2009

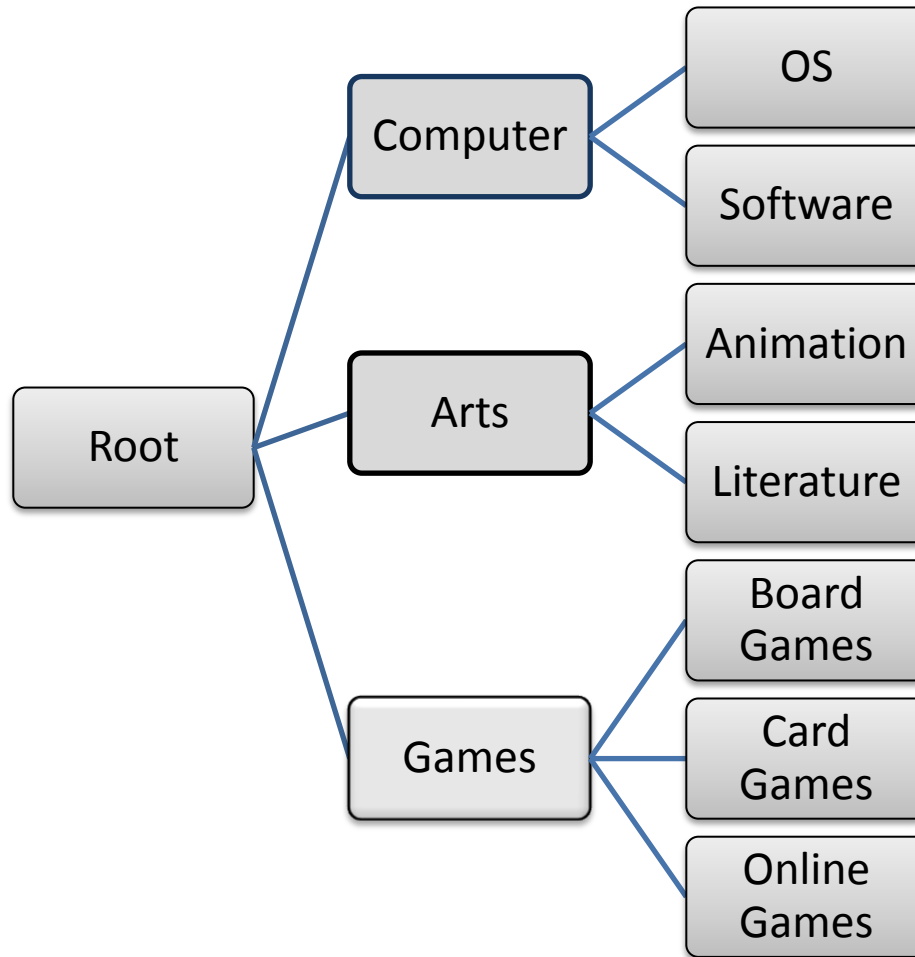
Paul Bennett, *Microsoft Research*

Joint work with Nam Nguyen, *Cornell University*

# Web Taxonomies

- Web taxonomies have been repeatedly demonstrated to be useful in improving browsing and search.  
[Dumais & Chen, 2000; Dumais *et al.*, 2001, Zhang *et al.*, 2005]
- Billions of web pages available versus millions that have been hand-categorized → need automatic classification methods.
- Web taxonomies with tens or hundreds of thousands of categories require approaches that can scale.
- Although many hierarchical algorithms, there is still significant need for improvement in hierarchical classification. [Liu *et al.*, 2005]

# An Example Taxonomy



# Classifying a Web Page



ኮህገግግ ገገገ  
*Creoso a'tel'Mithrim*

Welcome to the home of tel'Mithrim, or the Grey Company in the Common tongue. We are a guild made up of those with Elven blood who have banded together for good company and furthering Elven interests. We were formed in the city of Neverwinter, but have since moved on. Most of us now live in the village we call tel'Ruid, or the Glade, deep in the northern forests of Britannia.

ኮህገግግ ገገገ  
*i'Meneldalie*

*Lye duraser naa i'Meneldalie, tanya n'eldalie, imya tyara ri'avarien tyar, auruva ten' ai'er. Lye uma mani hye irma avaene awrien ai'er. Coia ulua im valina ar' malia ten' ai'er. Eller naa ai tanya ettula en'dagora, ar'hye avara dagor iire eller naa men. Nan'eller naa nat'or ten'man hye dagora biraetha. Tharien i'men en' nat'or n'sanuva yassen rangwien. Eldalie caela an rinar, dethola tyaralle yassen nomin pelu sen. Mellonea, ar'gothea, rinuva.*

Our most important belief is the Elven Way, that no elf, by action or inaction, shall cause harm to another elf. We do as we will as long as no harm is done. Life is a balance between happiness for yourself and concern for others. There is rarely a point to any conflict, and we try to avoid battles if we can. There are things, however, that we can not avoid. Disruption of the balance is not to be taken lightly. Elven memories are long ones, choose your actions wisely around them. Friends, and enemies, are not forgotten.



Copyright © 2000, The Grey Company. All Rights Reserved.  
Updated: 1/25/2002  
Send comments and questions to the [council@grey-company.org](mailto:council@grey-company.org)

# Classifying a Web Page



✦
✦

✦
✦

✦
✦

✦
✦

✦
✦

✦
✦

✦
✦

✦
✦

✦
✦

✦
✦

✦
✦

✦
✦

✦
✦

✦
✦

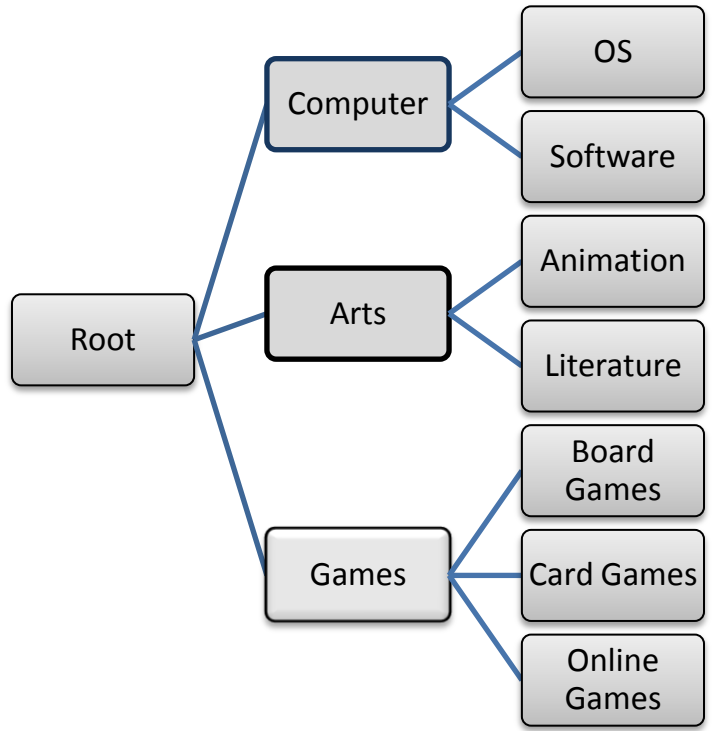
Welcome to the home of tel'Mithrim, or the Grey Company in the Common tongue. We are a guild made up of those with Elven blood who have banded together for good company and furthering Elven interests. We were formed in the city of Neverwinter, but have since moved on. Most of us now live in the village we call tel'Ruid, or the Glade, deep in the northern forests of Britannia.

*Lye duraser naa i'Menel.  
avara dagor iire eller naa rinuva.*

Our most important b  
for yourself and conce  
to be taken lightly. El'


*er. Eller naa ai tanya ettula en'dagora, ar'lye  
assen nomin pelu sen. Mellonea, ar'gothea,*

e. Life is a balance between happiness  
t avoid. Disruption of the balance is not



# Classifying a Web Page




 ቡ ላገጎ ገጠጠጎ  
*Creoso a'tel'Mithrim*

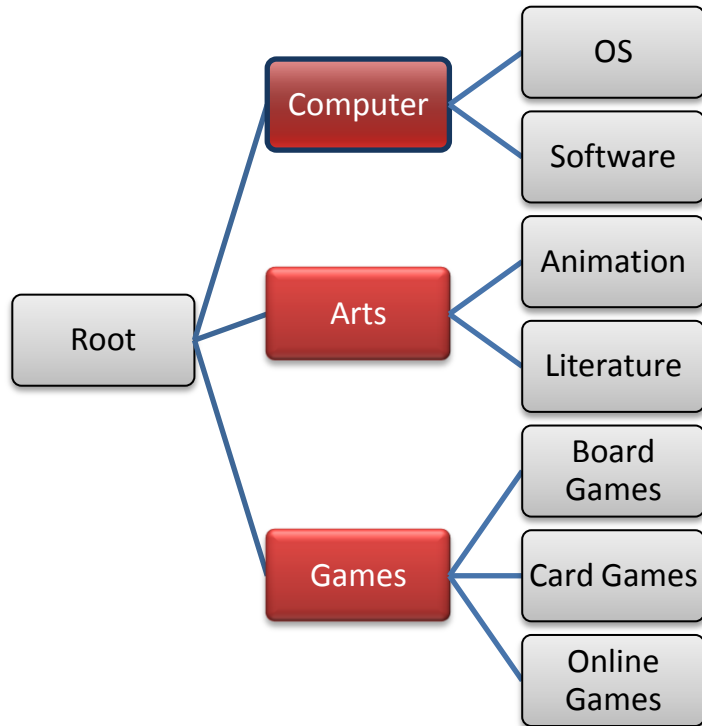
Welcome to the home of tel'Mithrim, or the Grey Company in the Common tongue. We are a guild made up of those with Elven blood who have banded together for good company and furthering Elven interests. We were formed in the city of Neverwinter, but have since moved on. Most of us now live in the village we call tel'Ruid, or the Glade, deep in the northern forests of Britannia.

*Lye duraser naa i'Menel.  
avara dagor iire eller naa rinuva.*

Our most important b  
for yourself and conce  
to be taken lightly. El'

*er. Eller naa ai tanya ettula en'dagora, ar'lye  
assen nomin pelu sen. Mellonea, ar'gothea,*

e. Life is a balance between happiness  
t avoid. Disruption of the balance is not



# Classifying a Web Page



✦
✦

✦
✦

✦
✦

✦
✦

✦
✦

✦
✦

✦
✦

✦
✦

✦
✦

✦
✦

✦
✦

✦
✦

✦
✦

✦
✦

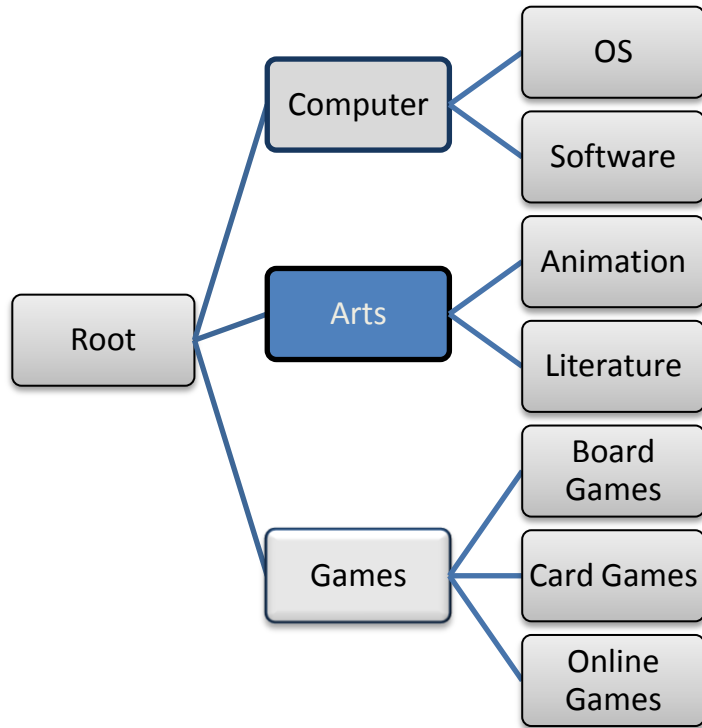
Welcome to the home of tel'Mithrim, or the Grey Company in the Common tongue. We are a guild made up of those with Elven blood who have banded together for good company and furthering Elven interests. We were formed in the city of Neverwinter, but have since moved on. Most of us now live in the village we call tel'Ruid, or the Glade, deep in the northern forests of Britannia.

*Lye duraser naa i'Menel.  
avara dagor iire eller naa rinuva.*

Our most important b  
for yourself and conce  
to be taken lightly. El'


*er. Eller naa ai tanya ettula en'dagora, ar'lye  
assen nomin pelu sen. Mellonea, ar'gothea,*

e. Life is a balance between happiness  
t avoid. Disruption of the balance is not



# Classifying a Web Page




  
 །་ ཡུཾ་ རྩཾ་ རྩཾ་
   
*Creoso a'tel'Mithrim*

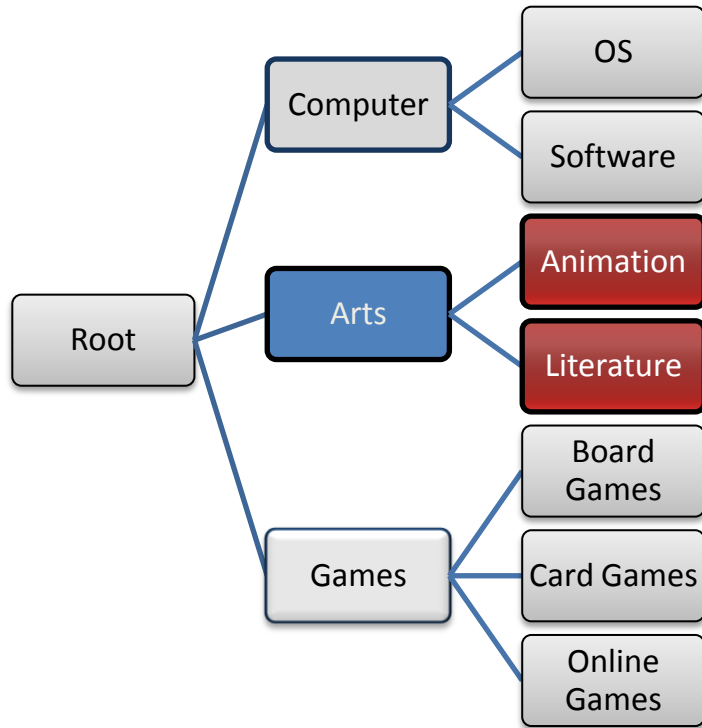
Welcome to the home of tel'Mithrim, or the Grey Company in the Common tongue. We are a guild made up of those with Elven blood who have banded together for good company and furthering Elven interests. We were formed in the city of Neverwinter, but have since moved on. Most of us now live in the village we call tel'Ruid, or the Glade, deep in the northern forests of Britannia.

*Lye duraser naa i'Menel.  
avara dagor iire eller naa rinuva.*

Our most important b  
for yourself and conce  
to be taken lightly. El'

*er. Eller naa ai tanya ettula en'dagora, ar'lye  
assen nomin pelu sen. Mellonea, ar'gothea,*

e. Life is a balance between happiness  
t avoid. Disruption of the balance is not



# Classifying a Web Page



✦
✦

✦
✦

✦
✦

✦
✦

✦
✦

✦
✦

✦
✦

✦
✦

✦
✦

✦
✦

✦
✦

✦
✦

✦
✦

✦
✦

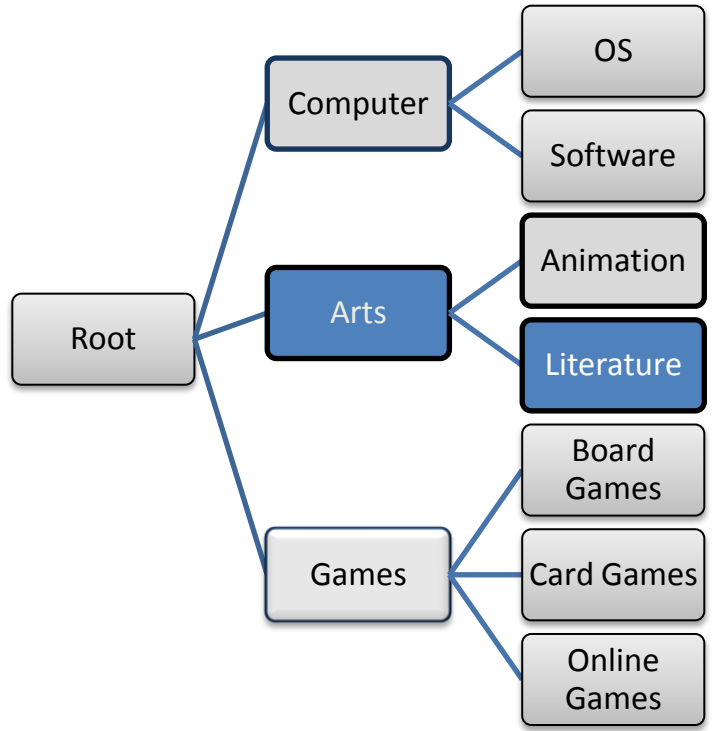
Welcome to the home of tel'Mithrim, or the Grey Company in the Common tongue. We are a guild made up of those with Elven blood who have banded together for good company and furthering Elven interests. We were formed in the city of Neverwinter, but have since moved on. Most of us now live in the village we call tel'Ruid, or the Glade, deep in the northern forests of Britannia.

*Lye duraser naa i'Menel, avara dagor iire eller naa rinuva.*

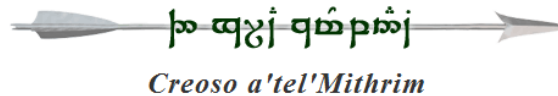
Our most important b for yourself and conce to be taken lightly. El'

*er. Eller naa ai tanya ettula en'dagora, ar'lye assen nomin pelu sen. Mellonea, ar'gothea,*

e. Life is a balance between happiness t avoid. Disruption of the balance is not



# The Actual Label?



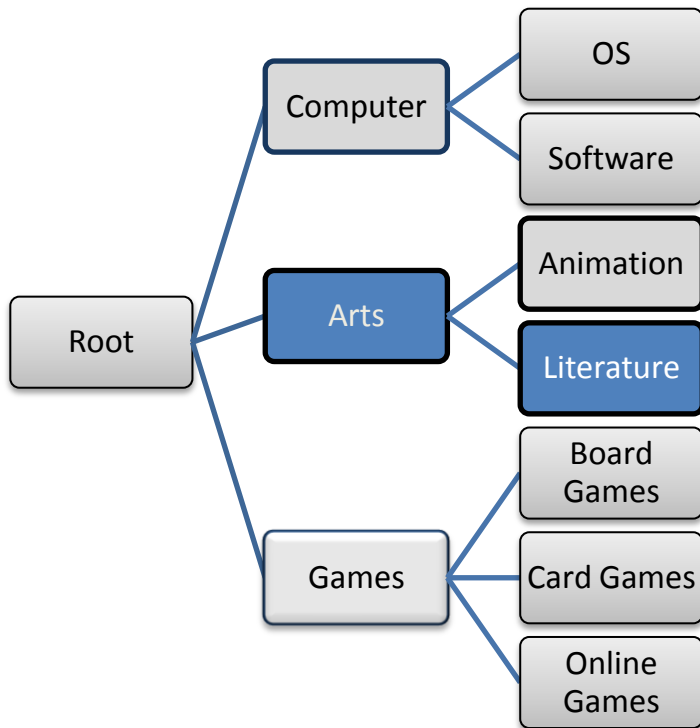
Welcome to the home of tel'Mithrim, or the Grey Company in the Common tongue. We are a guild made up of those with Elven blood who have banded together for good company and furthering Elven interests. We were formed in the city of Neverwinter, but have since moved on. Most of us now live in the village we call tel'Ruid, or the Glade, deep in the northern forests of Britannia.

*Lye duraser naa i'Menel.  
avara dagor iire eller naa rinuva.*

Our most important b  
for yourself and conce  
to be taken lightly. El'

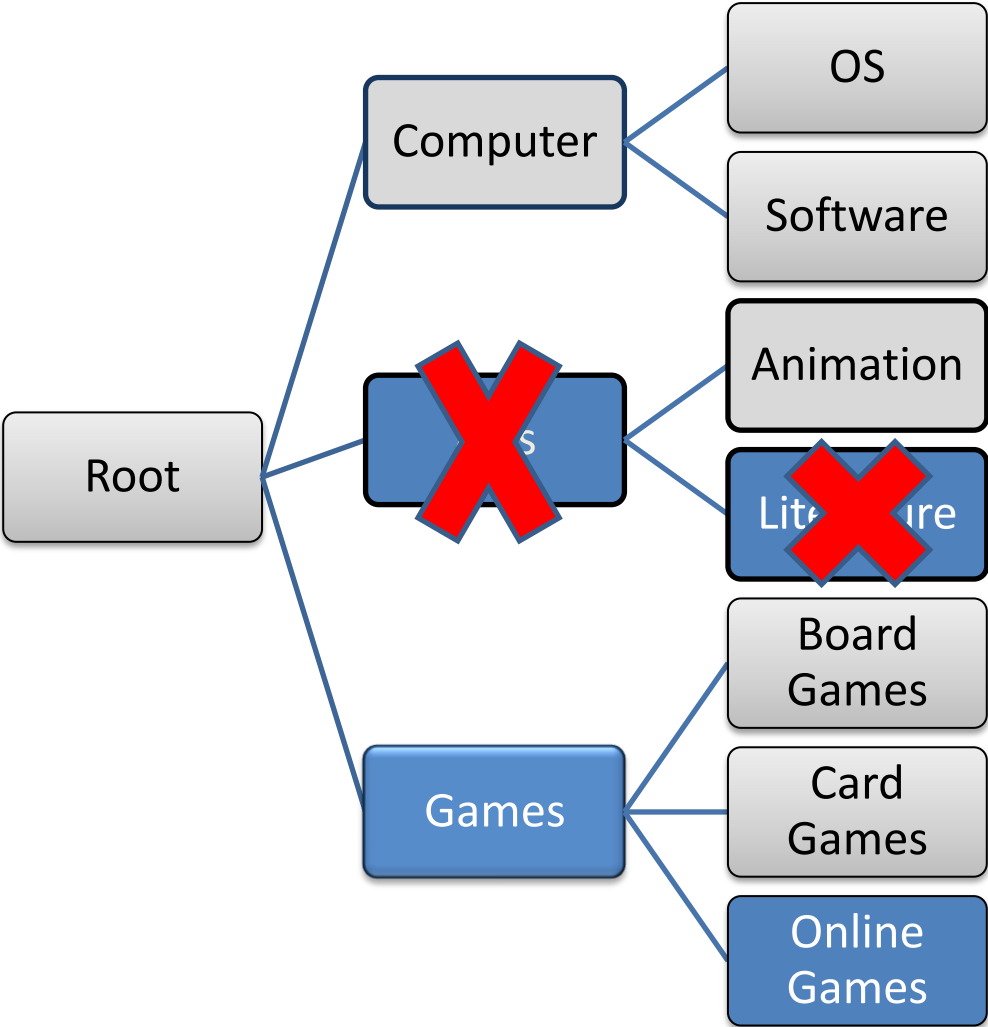
*er. Eller naa ai tanya ettula en'dagora, ar'lye  
assen nomin pelu sen. Mellonea, ar'gothea,*

e. Life is a balance between happiness  
t avoid. Disruption of the balance is not

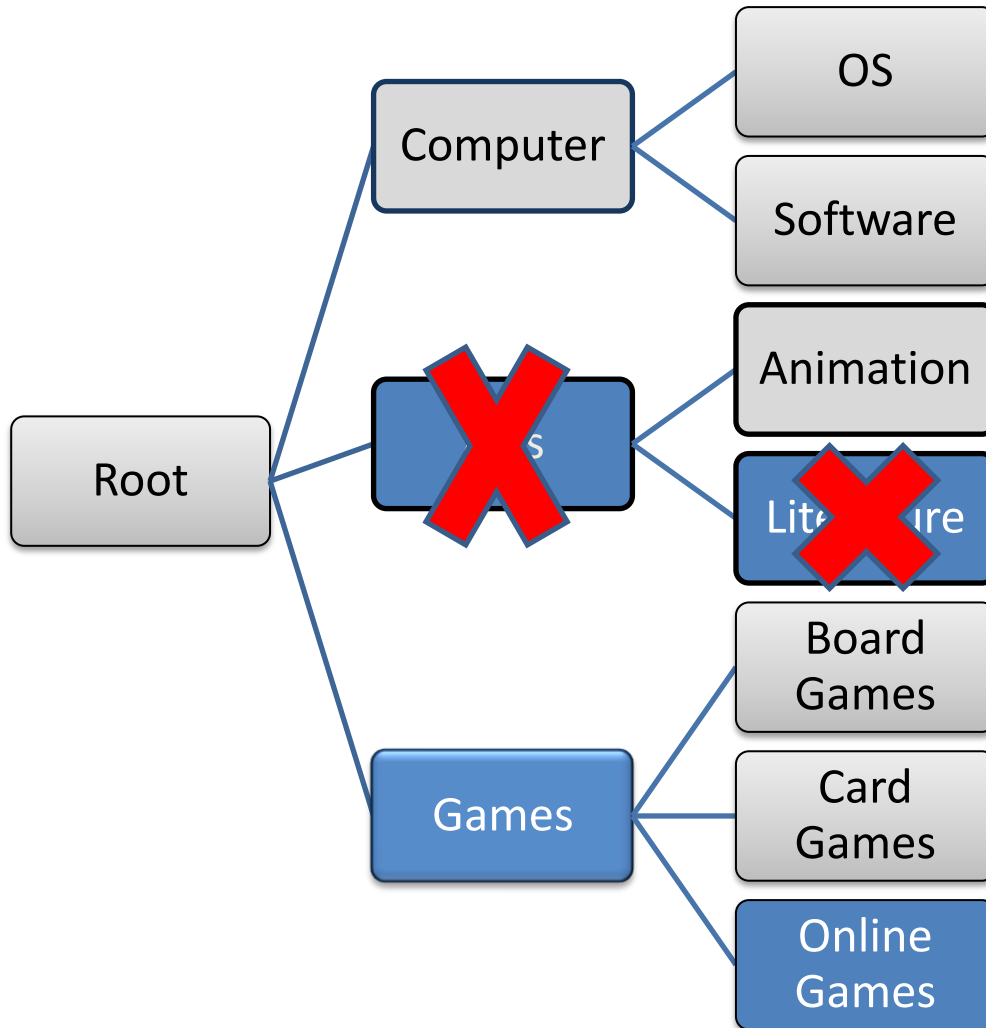




# Hierarchical Classification Challenges



# Hierarchical Classification Challenges



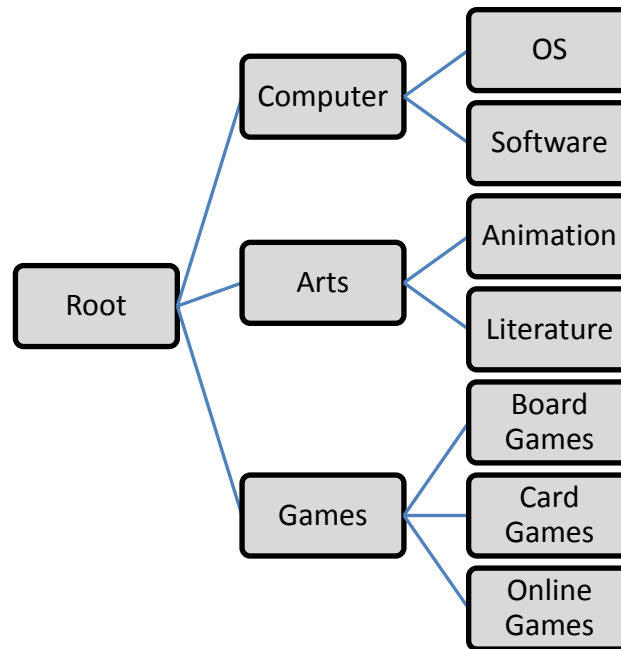
Error propagation

- An error made higher up cannot be corrected.

Complex decision surfaces **high** in the hierarchy

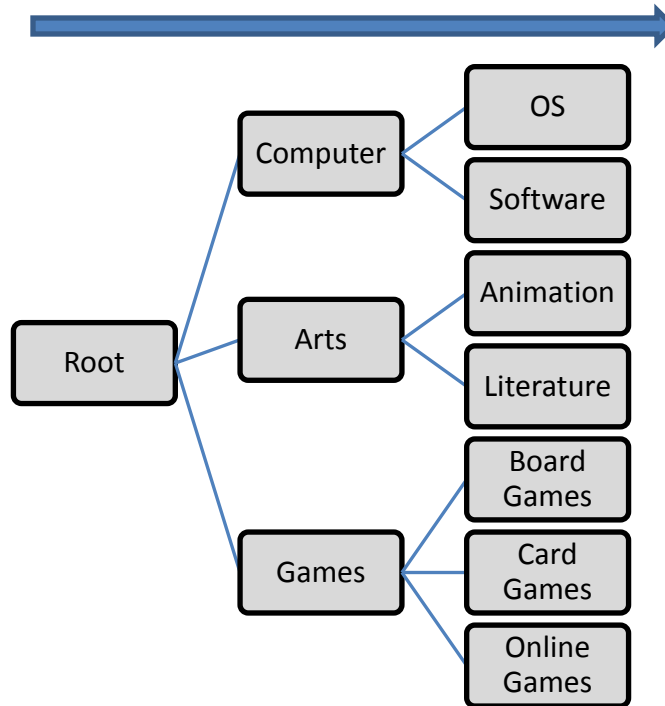
- Meta-reviewers as upper nodes and reviewers as leaves.
- Ask reviewers to read *abstract* and predict before assigning to meta.

# Refinement



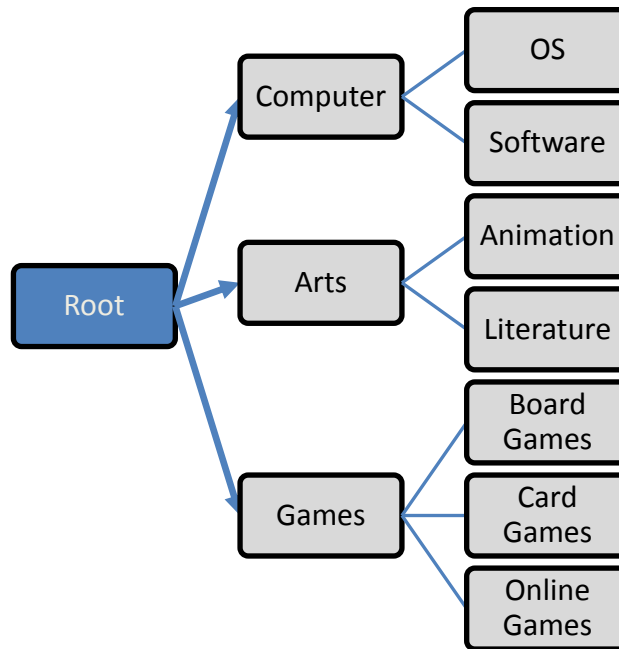
# Refinement

Top-down to reduce error propagation



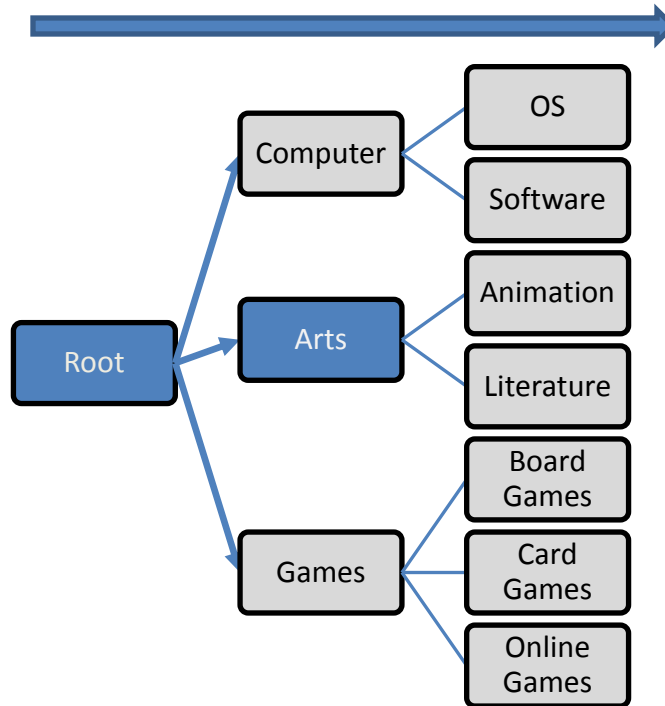
# Refinement

Top-down to reduce error propagation



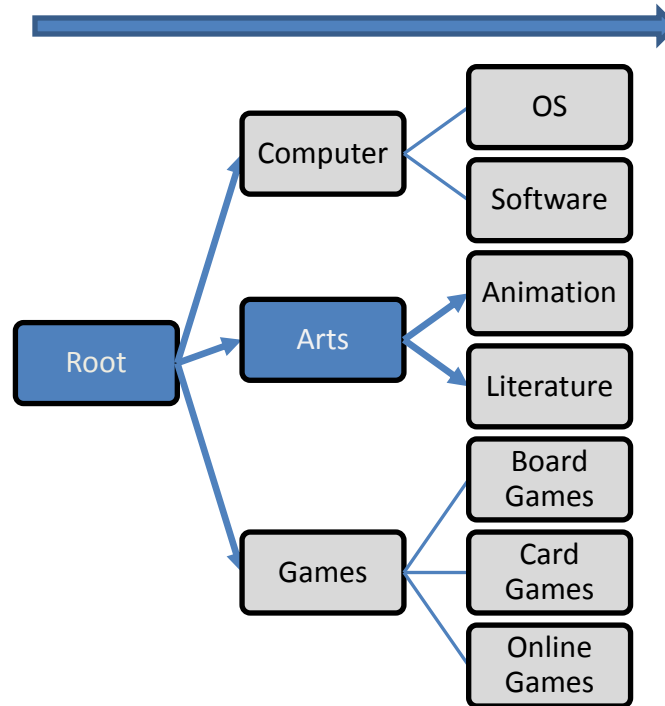
# Refinement

Top-down to reduce error propagation



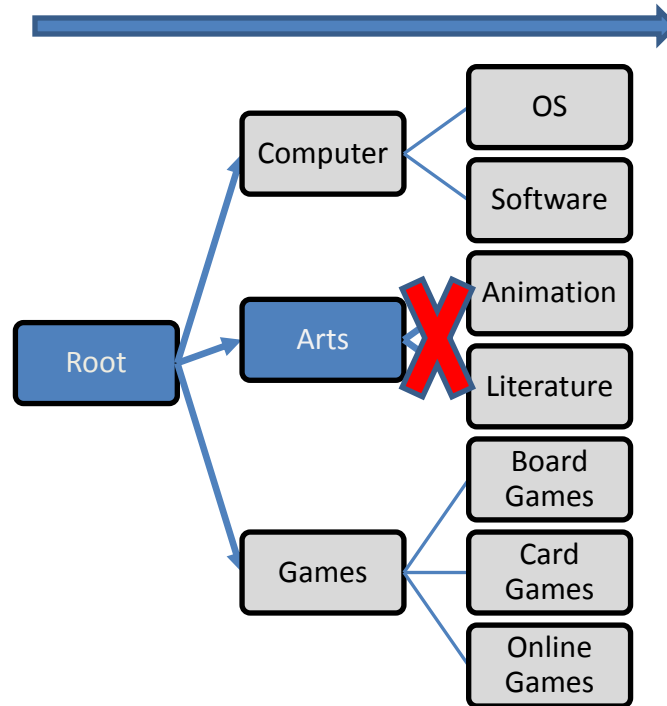
# Refinement

Top-down to reduce error propagation

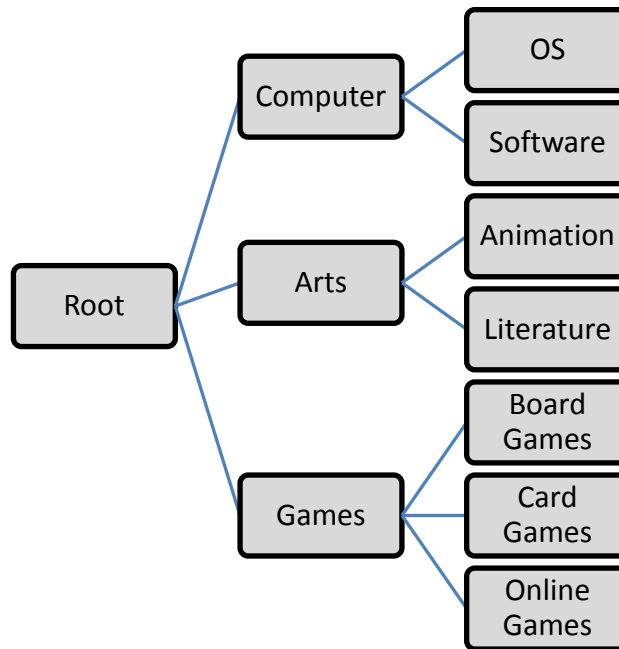


# Refinement

Top-down to reduce error propagation

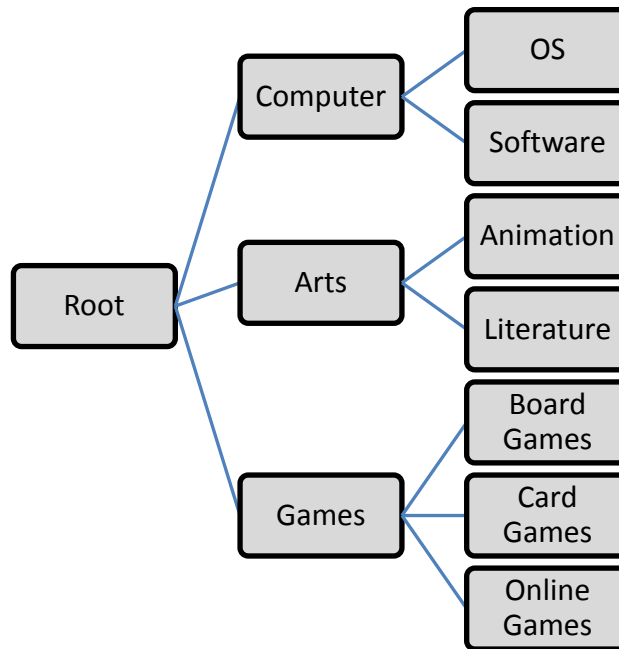


# Refined Experts



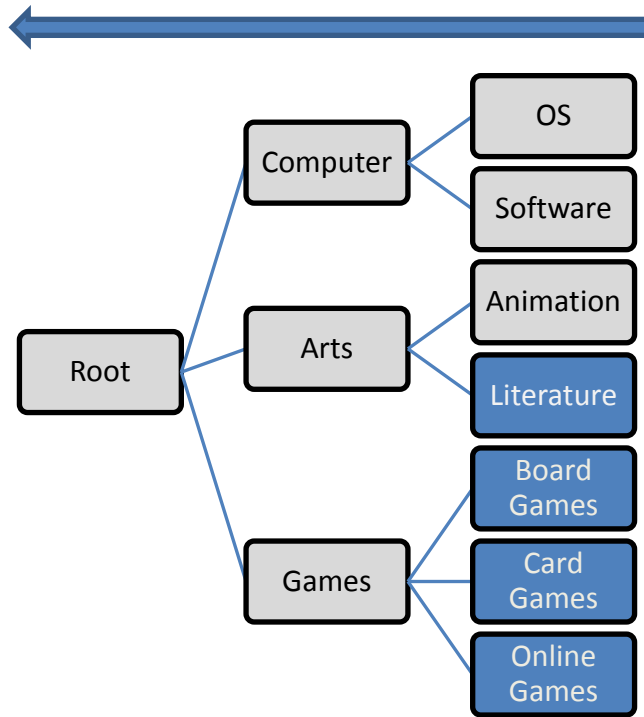
# Refined Experts

Bottom-up to propagate up “first-guess” expert information



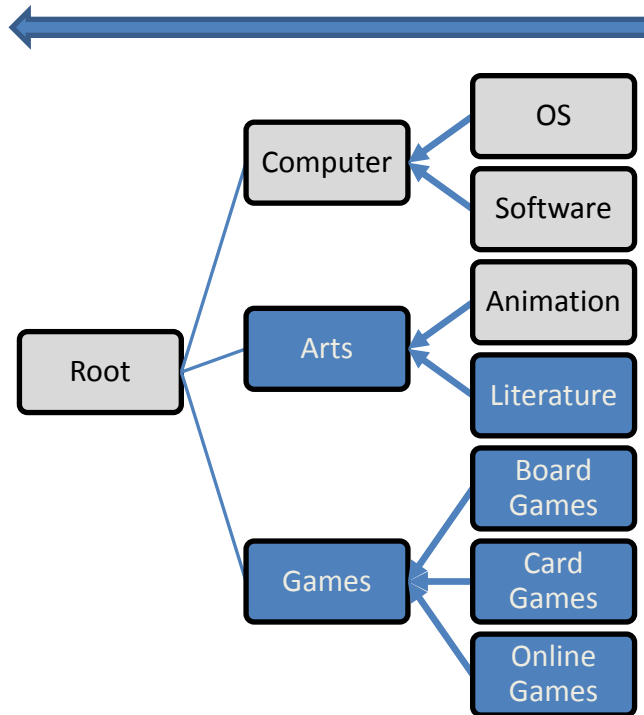
# Refined Experts

Bottom-up to propagate up “first-guess” expert information



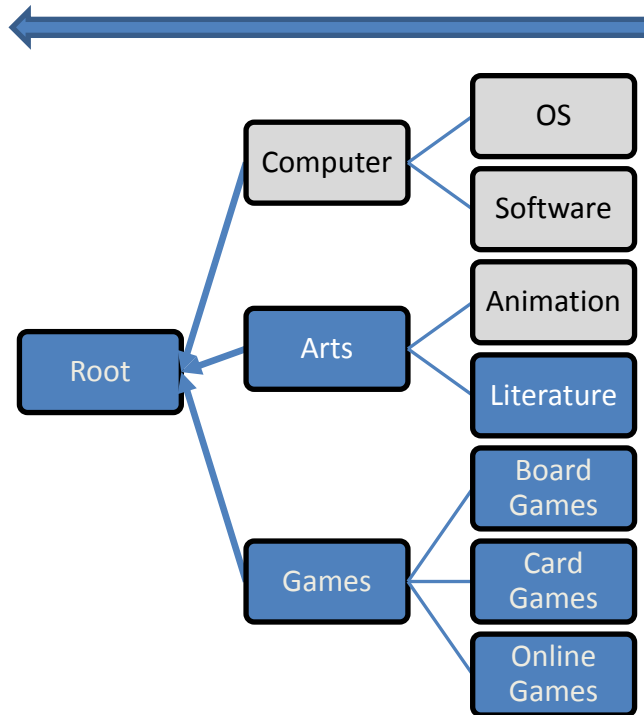
# Refined Experts

Bottom-up to propagate up “first-guess” expert information



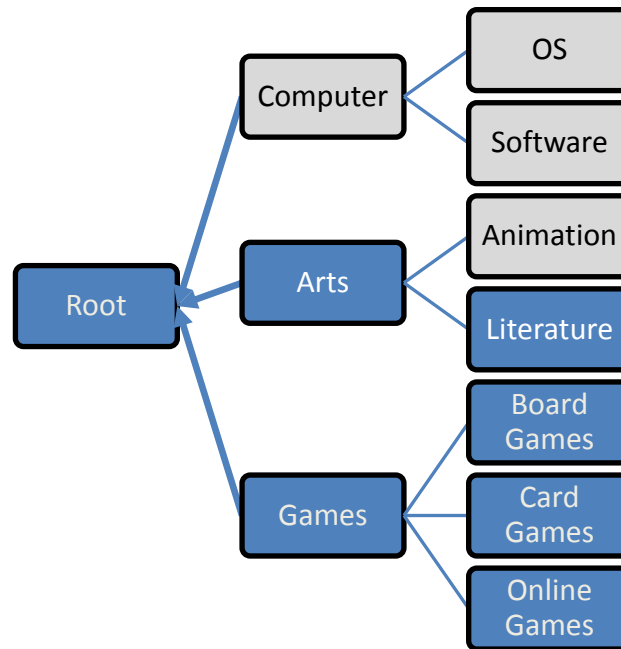
# Refined Experts

Bottom-up to propagate up “first-guess” expert information



# Refined Experts

Bottom-up to propagate up “first-guess” expert information

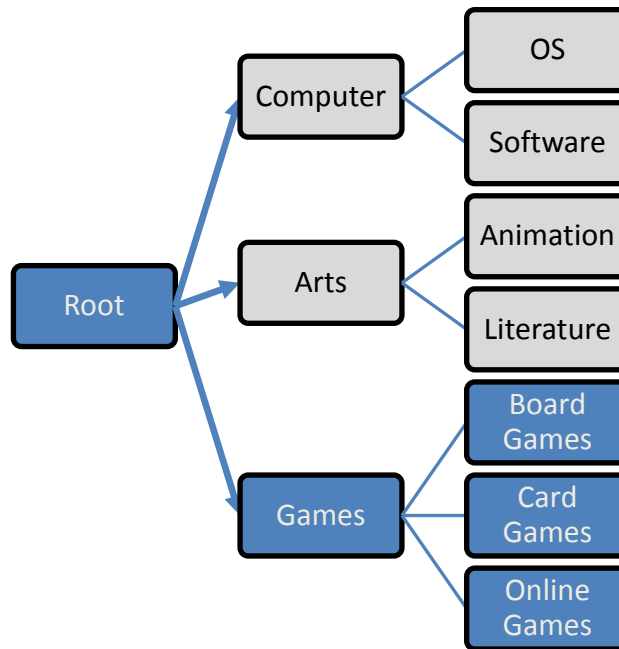


Top-down to reduce error propagation



# Refined Experts

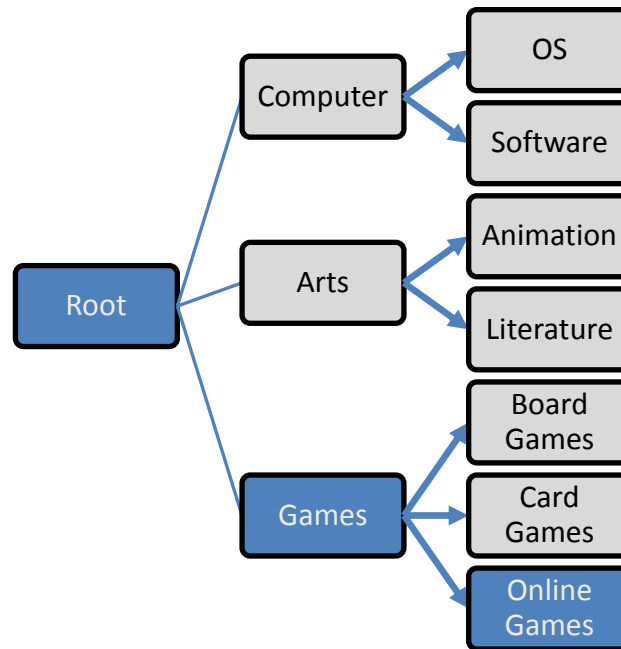
Bottom-up to propagate up “first-guess” expert information



Top-down to reduce error propagation

# Refined Experts

Bottom-up to propagate up “first-guess” expert information

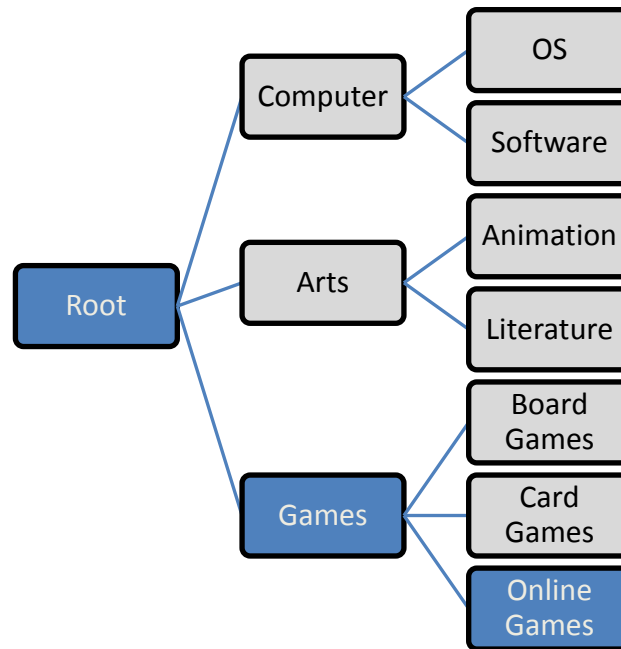


Top-down to reduce error propagation



# Refined Experts

Bottom-up to propagate up “first-guess” expert information



Top-down to reduce error propagation

# Problems with Hierarchical Classification

# Problems with Hierarchical Classification

1. Data are very sparse at the leaf nodes in a hierarchy.

# Problems with Hierarchical Classification

1. Data are very sparse at the leaf nodes in a hierarchy.
2. Errors made at higher levels propagate to lower levels. → **predict errors and stop propagation.**

# Problems with Hierarchical Classification

1. Data are very sparse at the leaf nodes in a hierarchy.
2. Errors made at higher levels propagate to lower levels. → **predict errors and stop propagation.**
3. As we travel up the hierarchy, the diffusivity of general concepts can lead to more complex decision surfaces. → **non-linear models.**

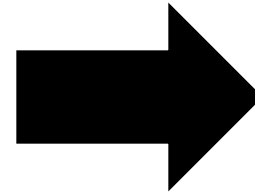
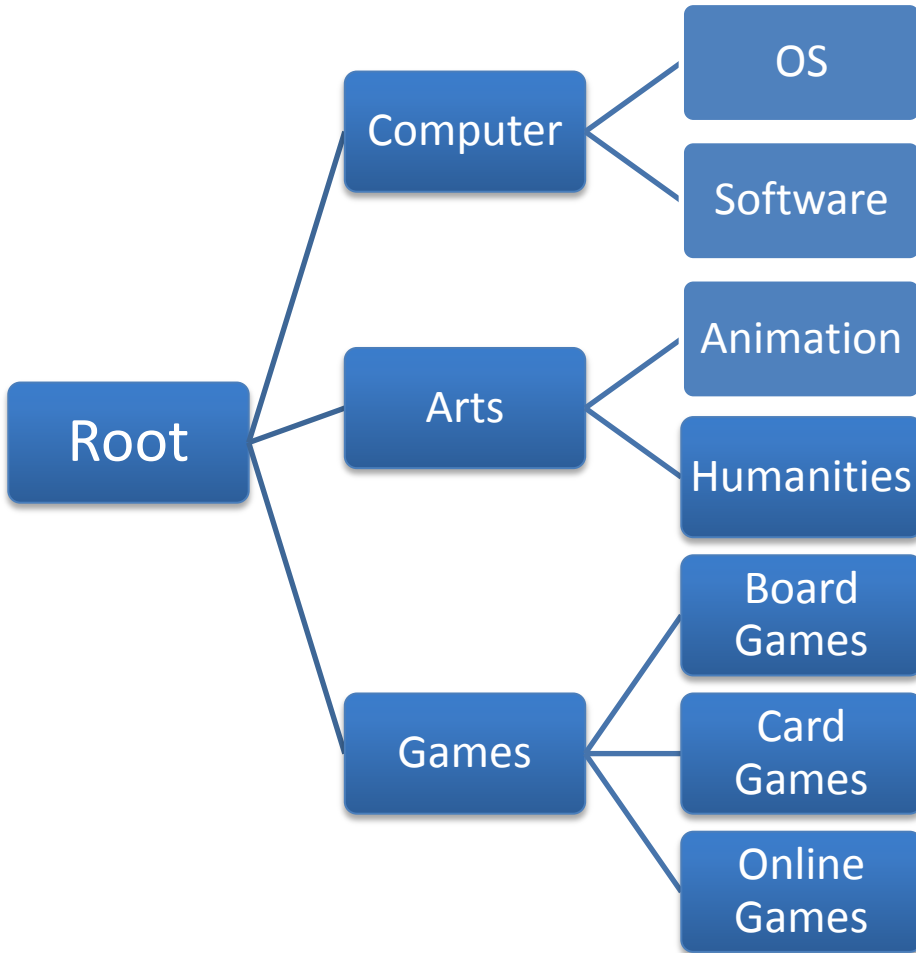
# Problems with Hierarchical Classification

1. Data are very sparse at the leaf nodes in a hierarchy.
2. Errors made at higher levels propagate to lower levels. → **predict errors and stop propagation.**
3. As we travel up the hierarchy, the diffusivity of general concepts can lead to more complex decision surfaces. → **non-linear models.**

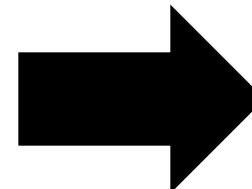
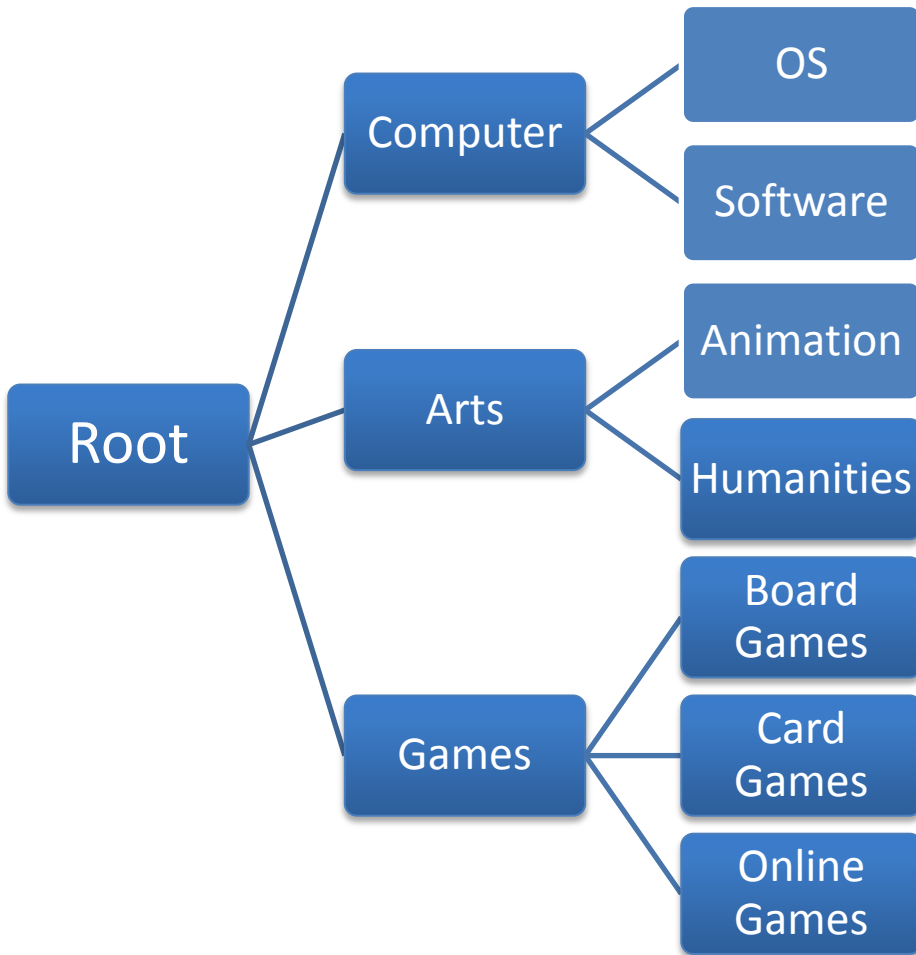
# Outline

1. Overview of Approach
2. Previous Work
3. Refinement & Refined Experts
4. Experiments and Results

# Flat Learner

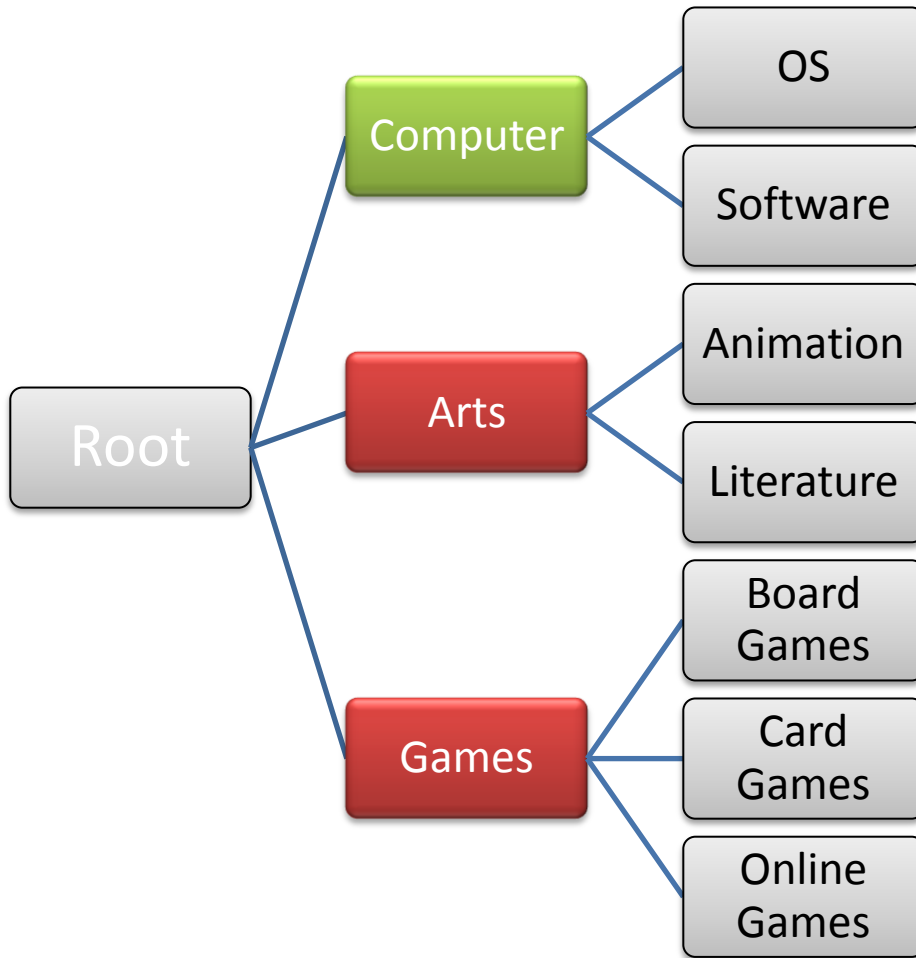


# Flat Learner

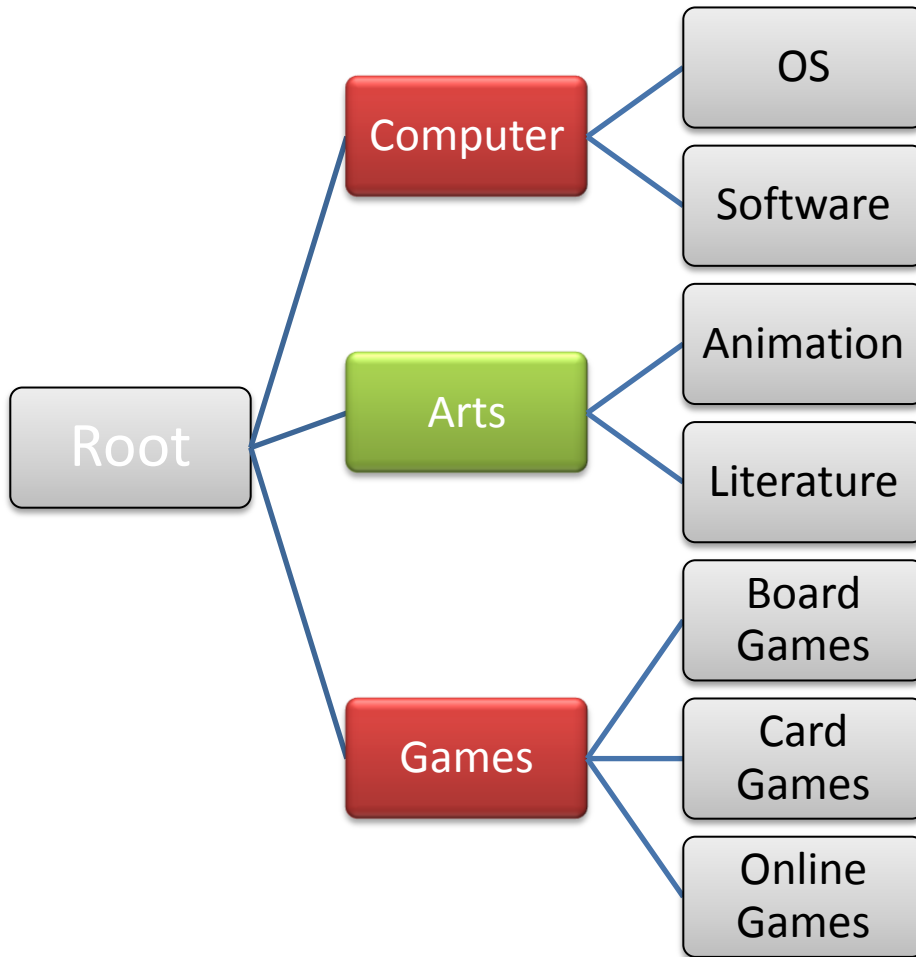


Training data for each category is the entire data set.

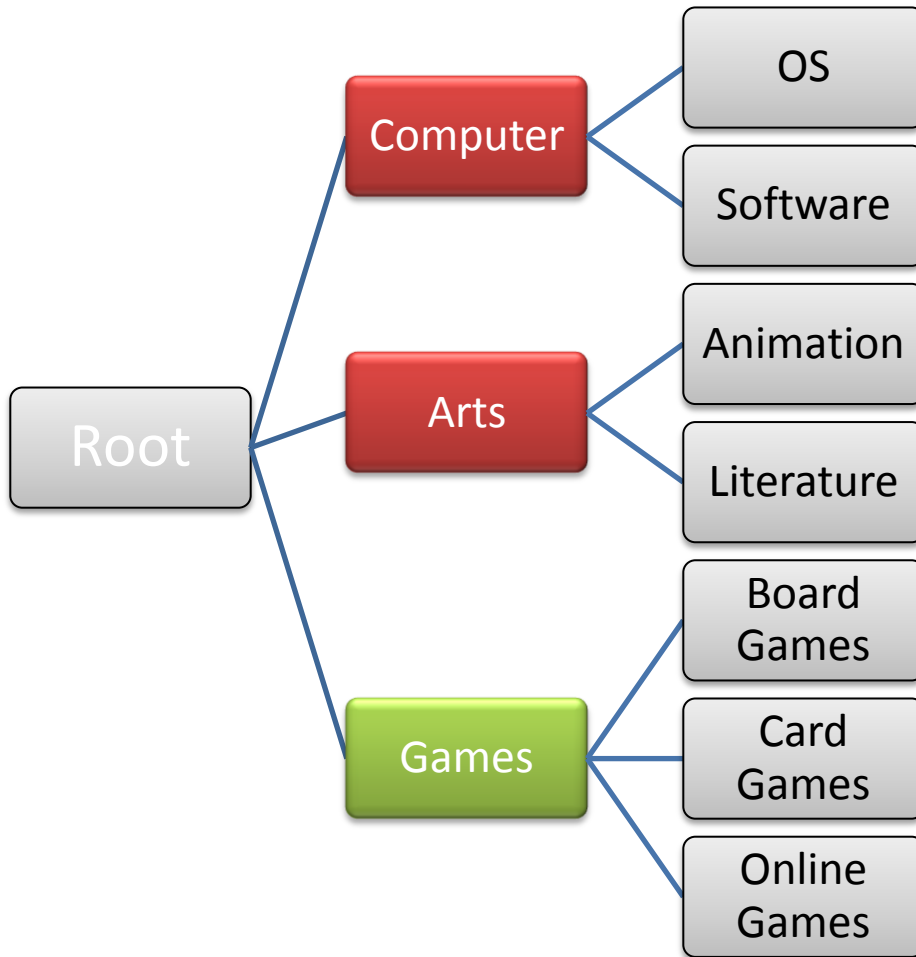
# Hierarchical Learner



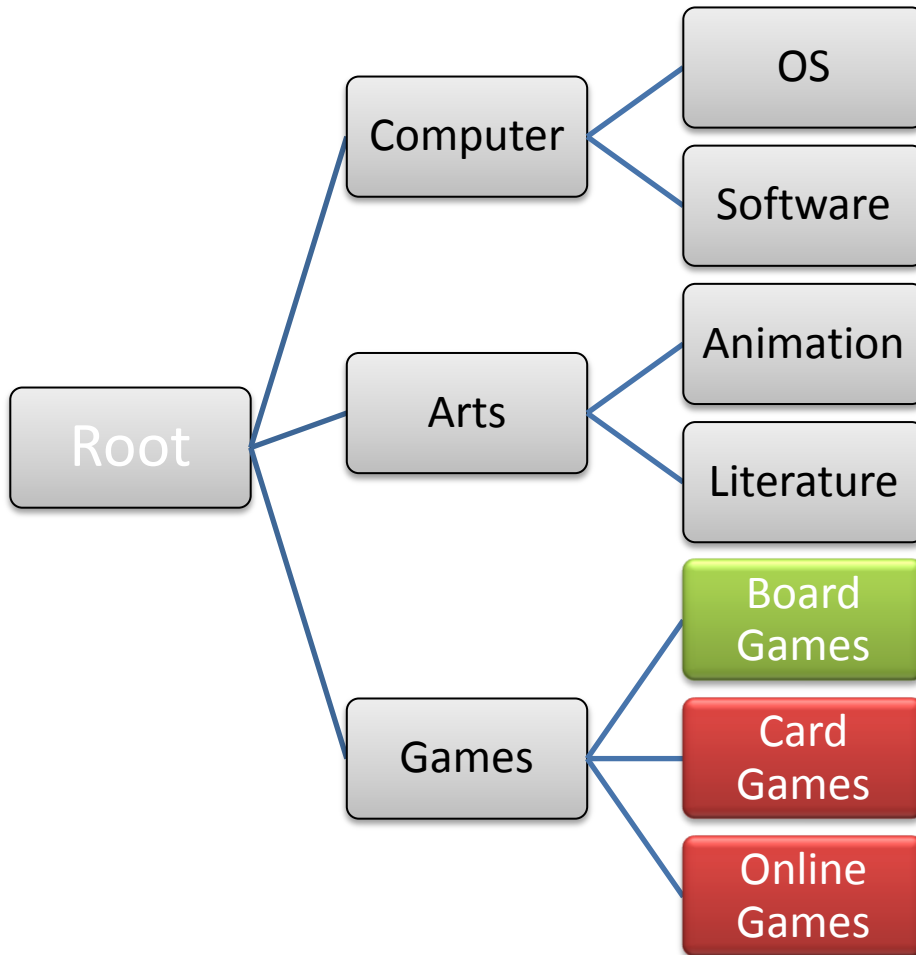
# Hierarchical Learner



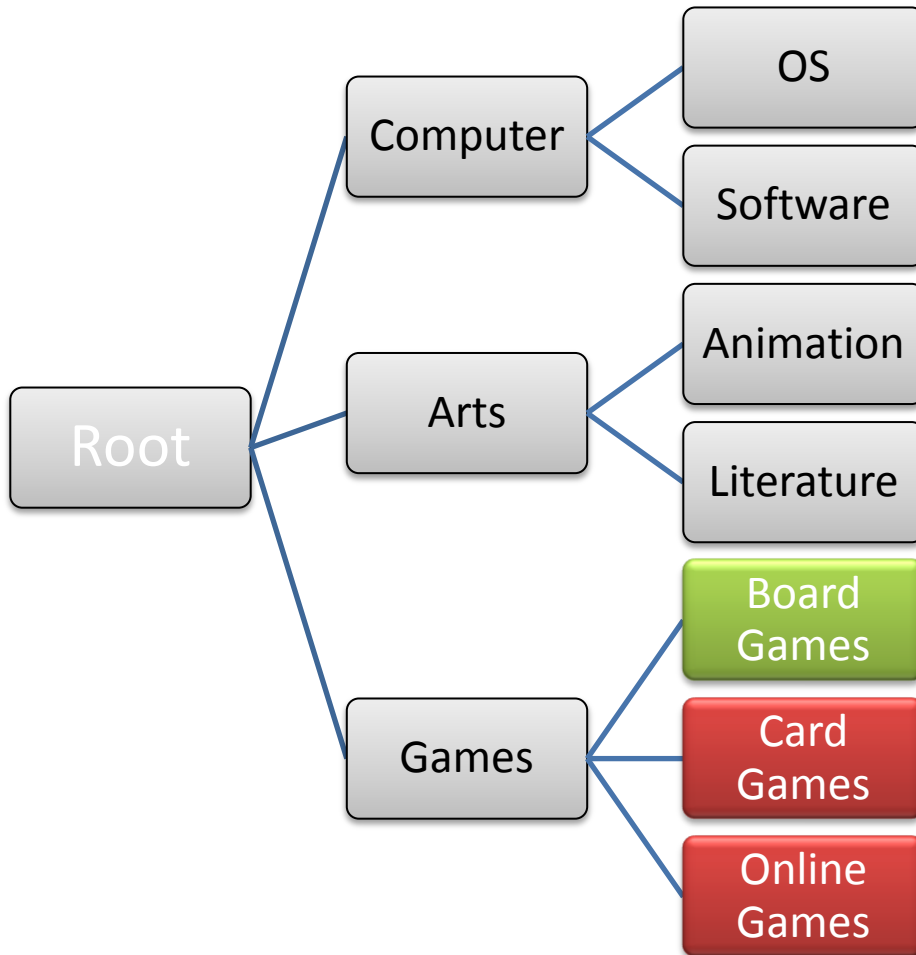
# Hierarchical Learner



# Hierarchical Learner

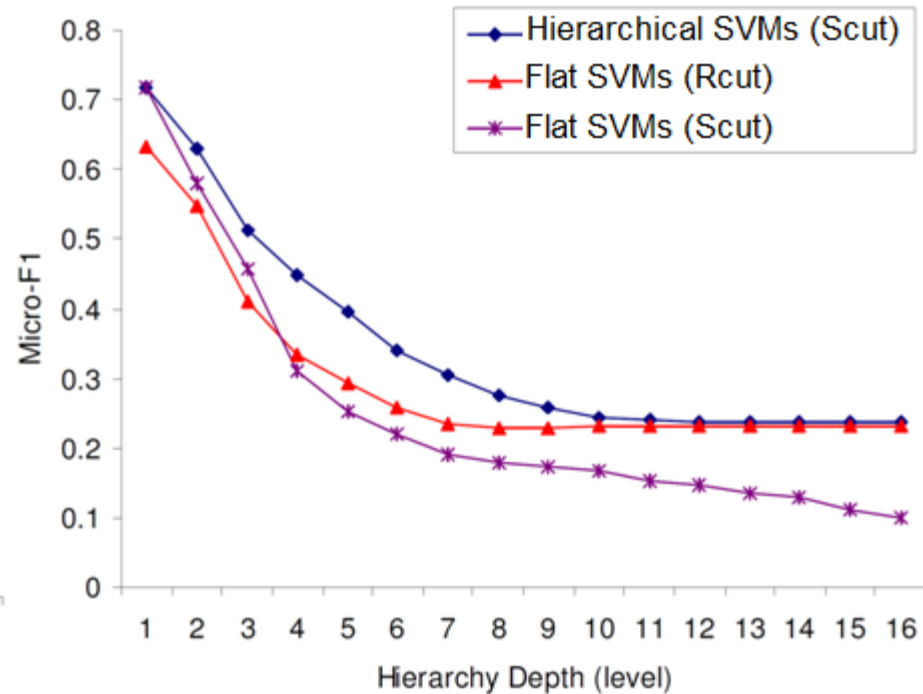
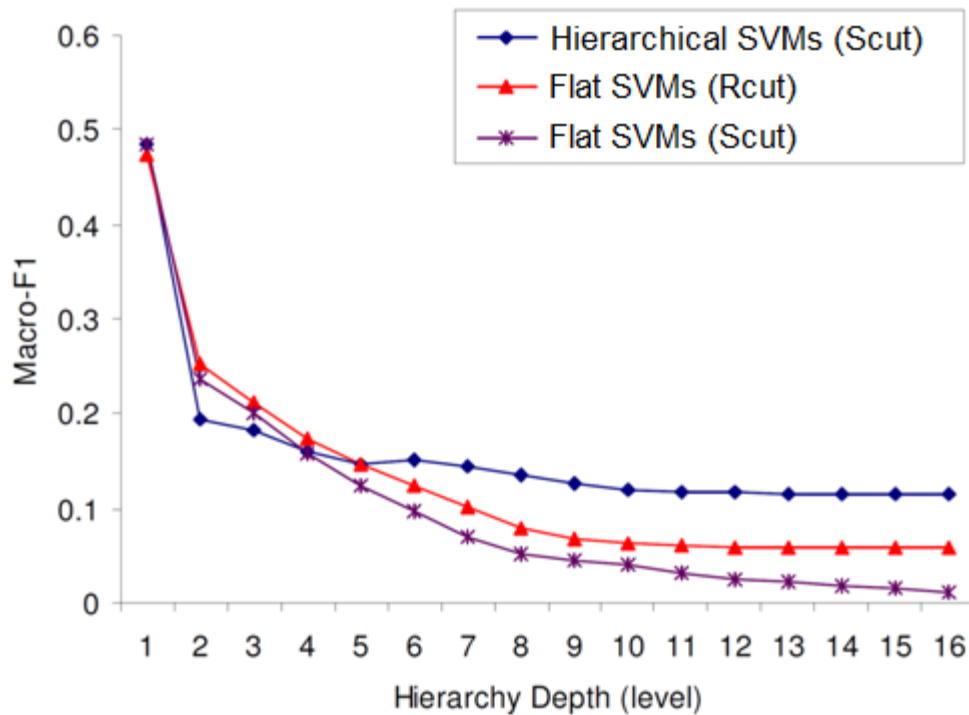


# Hierarchical Learner



The training examples of the **board games** category include all documents belonging to the **games** categories.

# Hierarchical Outperforms Flat



Figures are taken from “Support Vector Machines Classification with A Very Large-scale Taxonomy”, Liu et. al. 2005.

# Other Select Related Work

- McCallum *et al.* (1998)
  - Shrinkage of parameters based on parents to help improve data sparsity problems.
- Jordan & Jacobs (1994) HME, Ruiz & Srinivasan (1999) variant of HME
  - Experts at leaves and gating at internal nodes fit with EM algorithm.
- Xue *et al.* (2008)
  - Two stage – category search to reduce to a smaller dynamic hierarchy and then classify into smaller hierarchy.

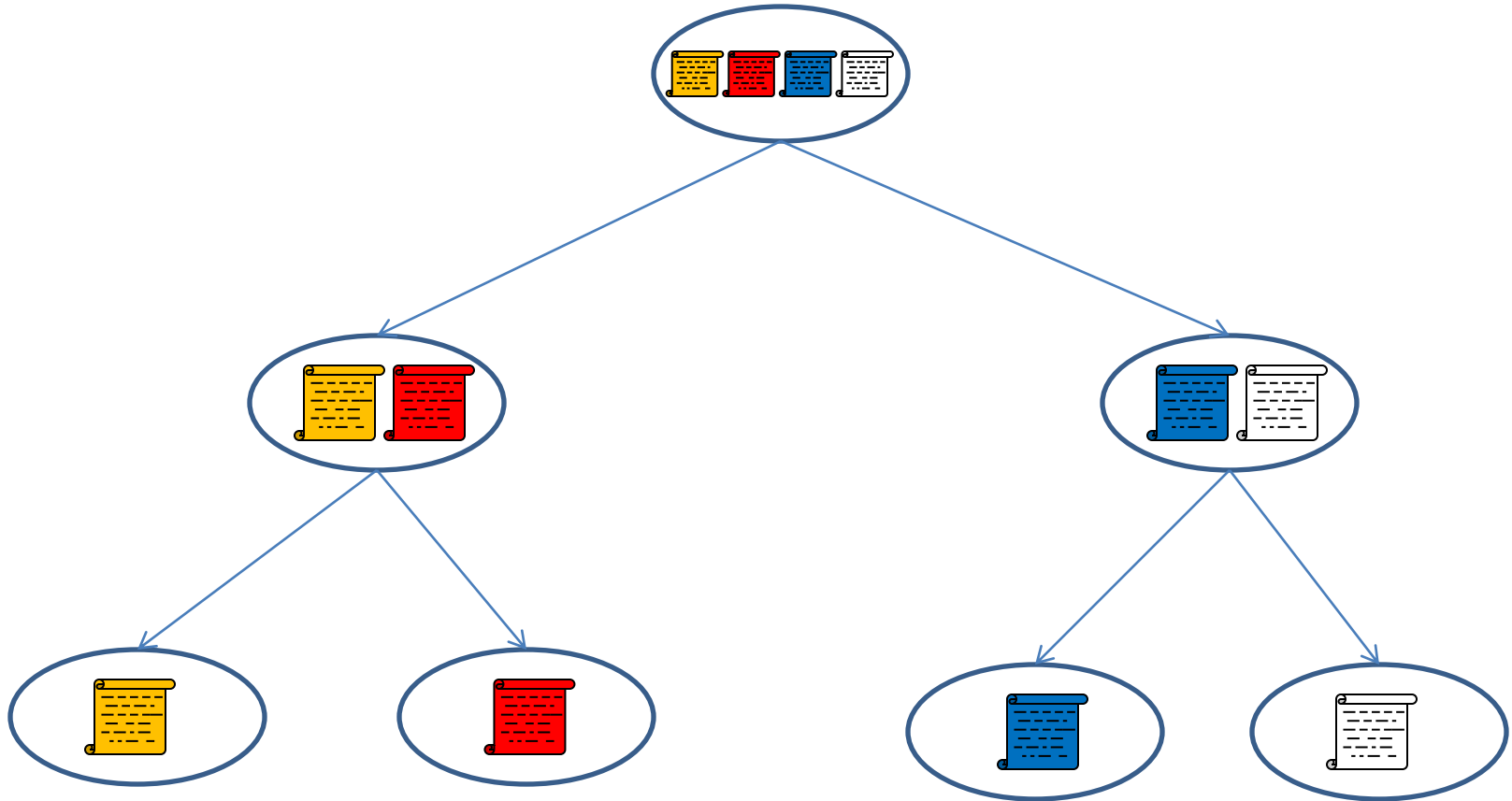
# Our Focus

- Identify and address two underlying machine learning issues for hierarchical classification
- Derive general framework applicable to other base learning algorithms
- Focus on large-scale taxonomies where joint optimization is difficult

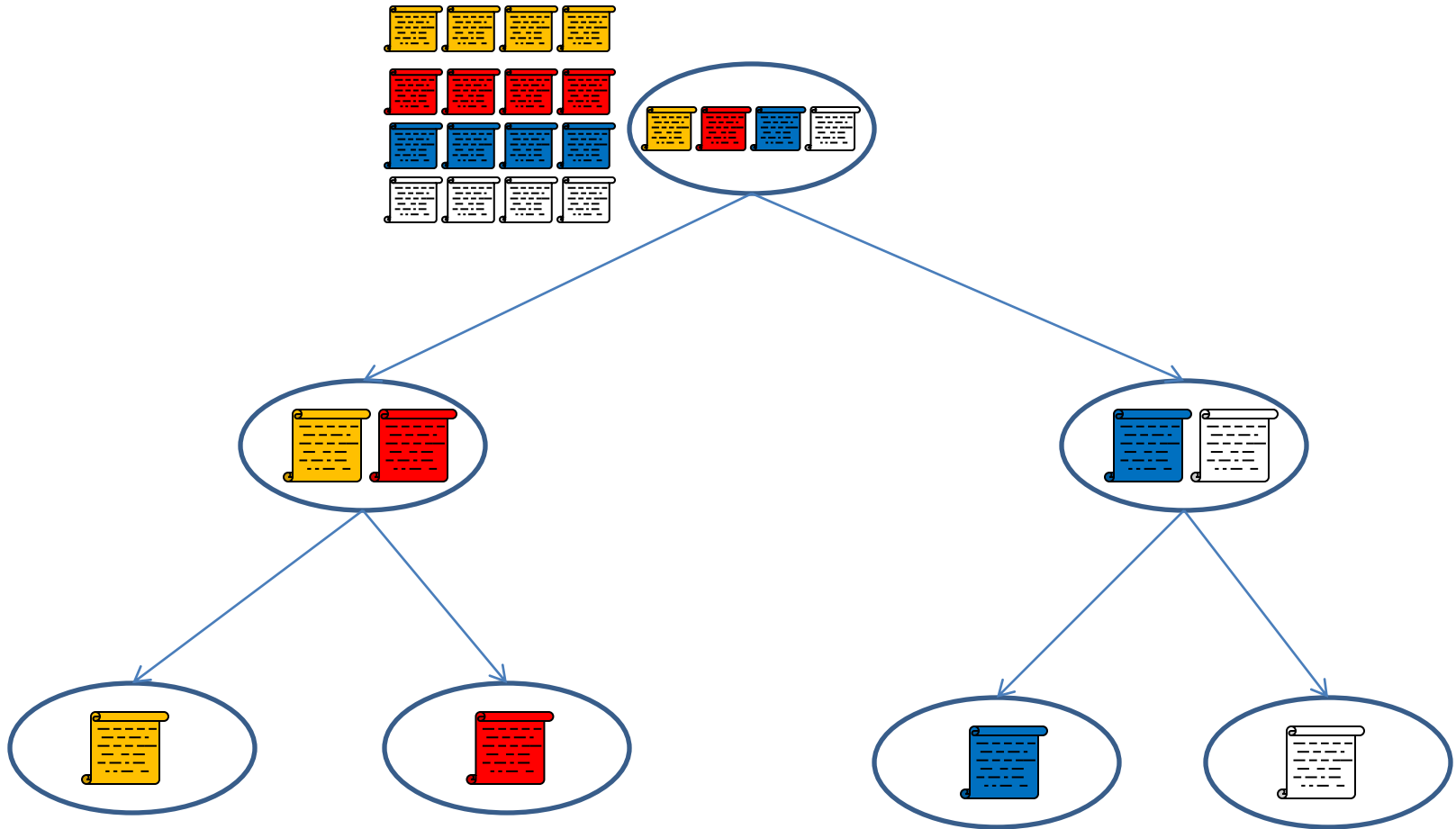
# Outline

1. Overview of Approach
2. Previous Work
3. Refinement & Refined Experts
4. Experiments and Results

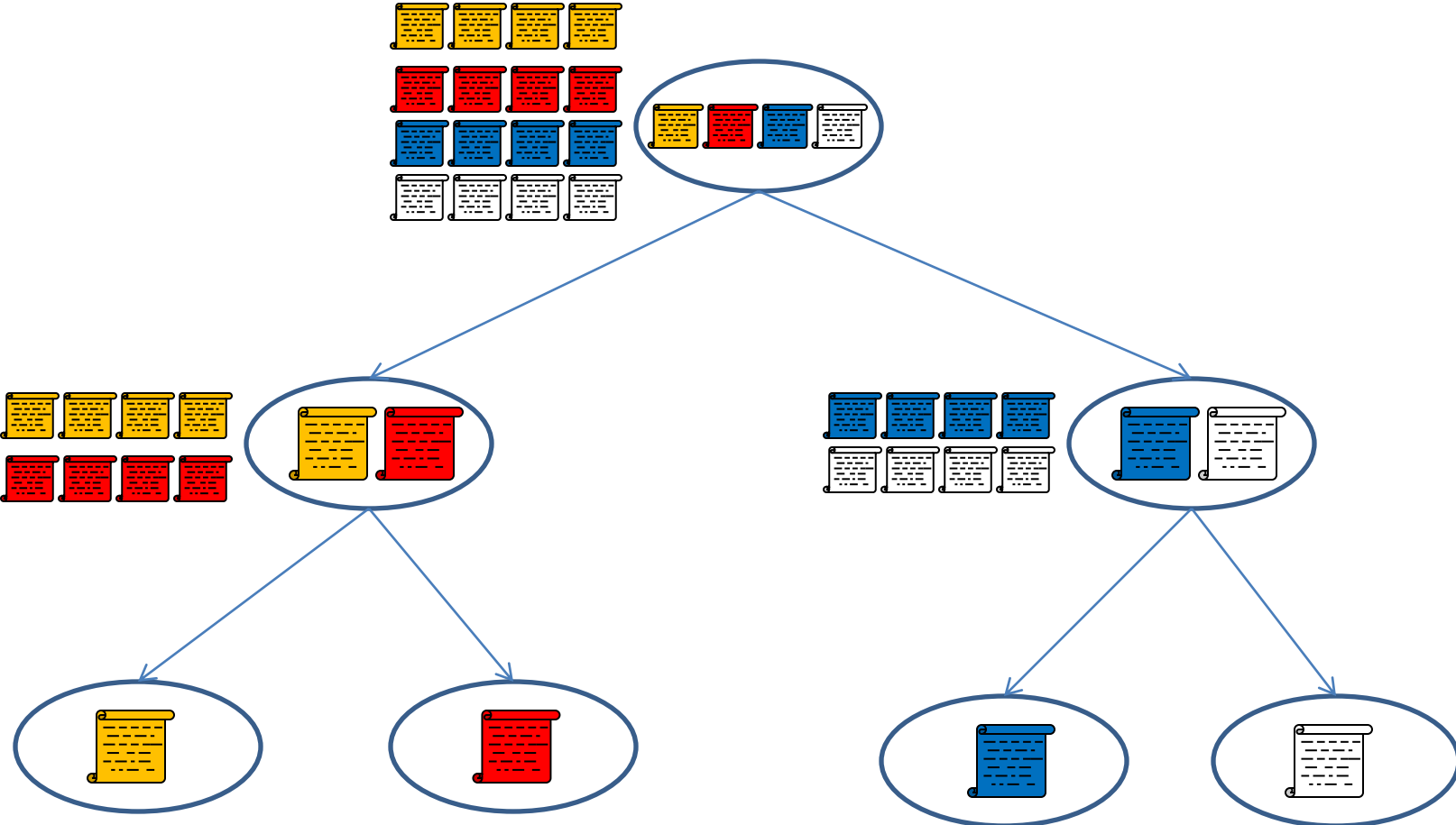
# Test Distribution Drift Example



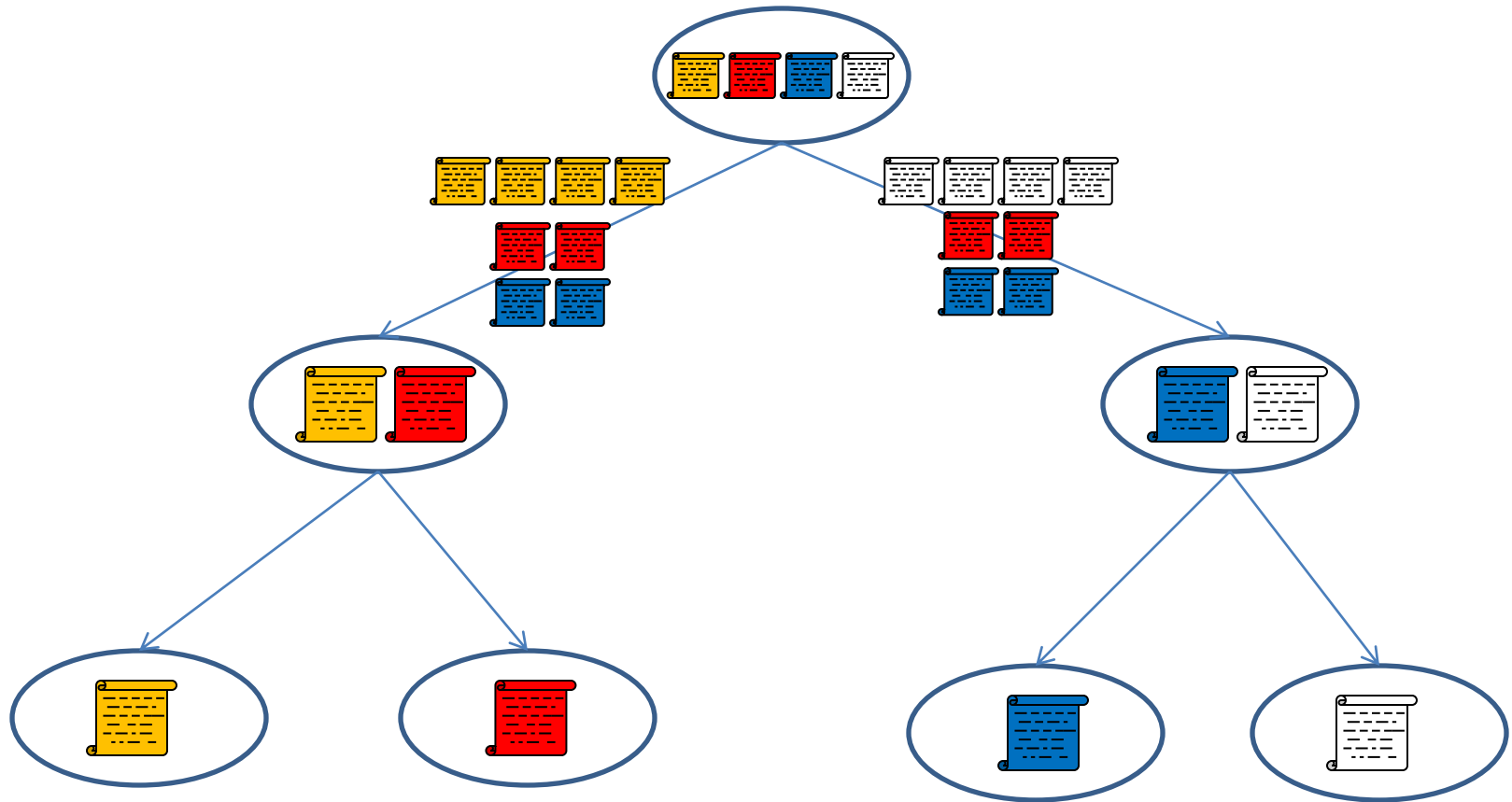
# Training Distribution



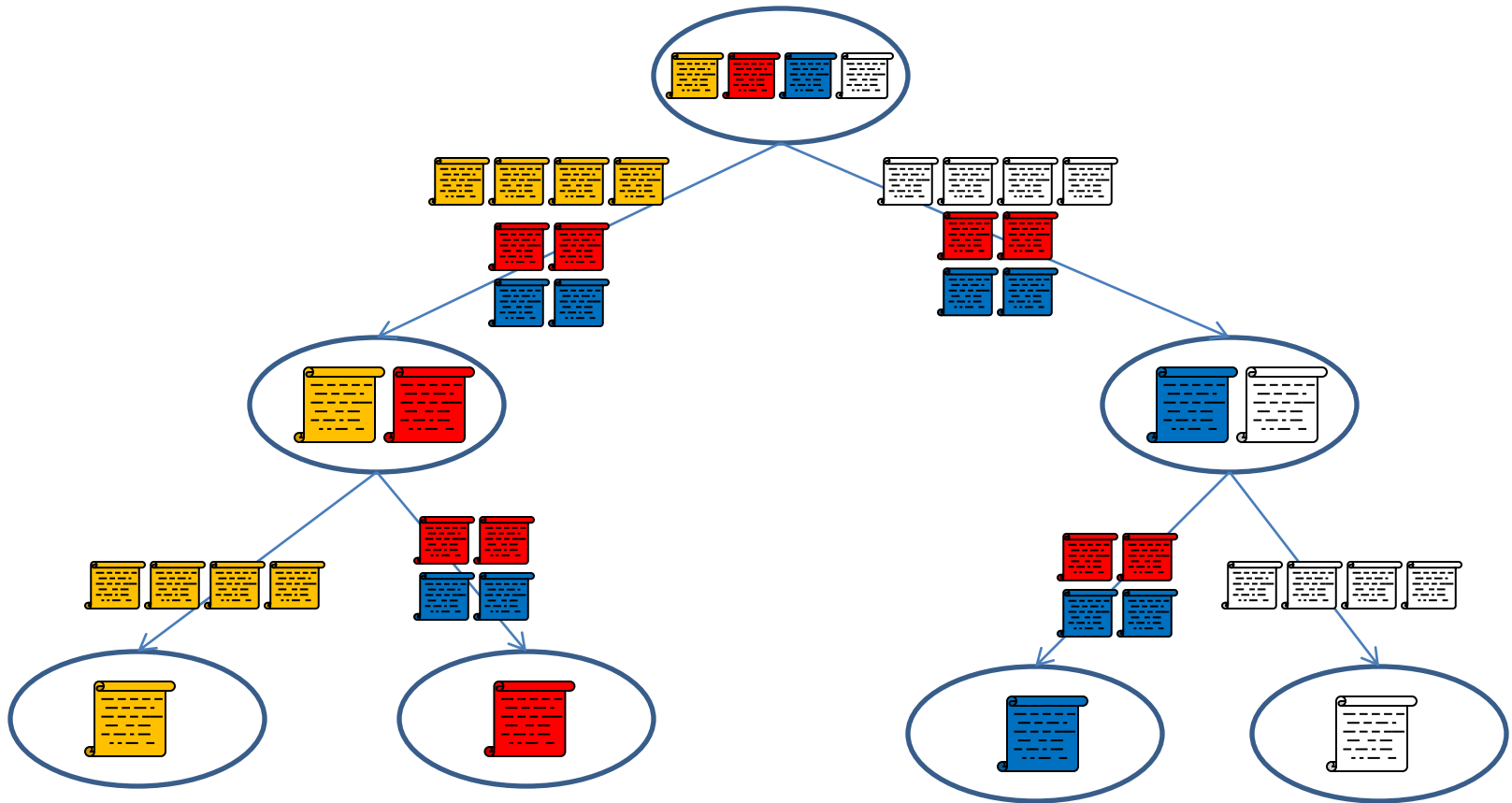
# Training Distribution



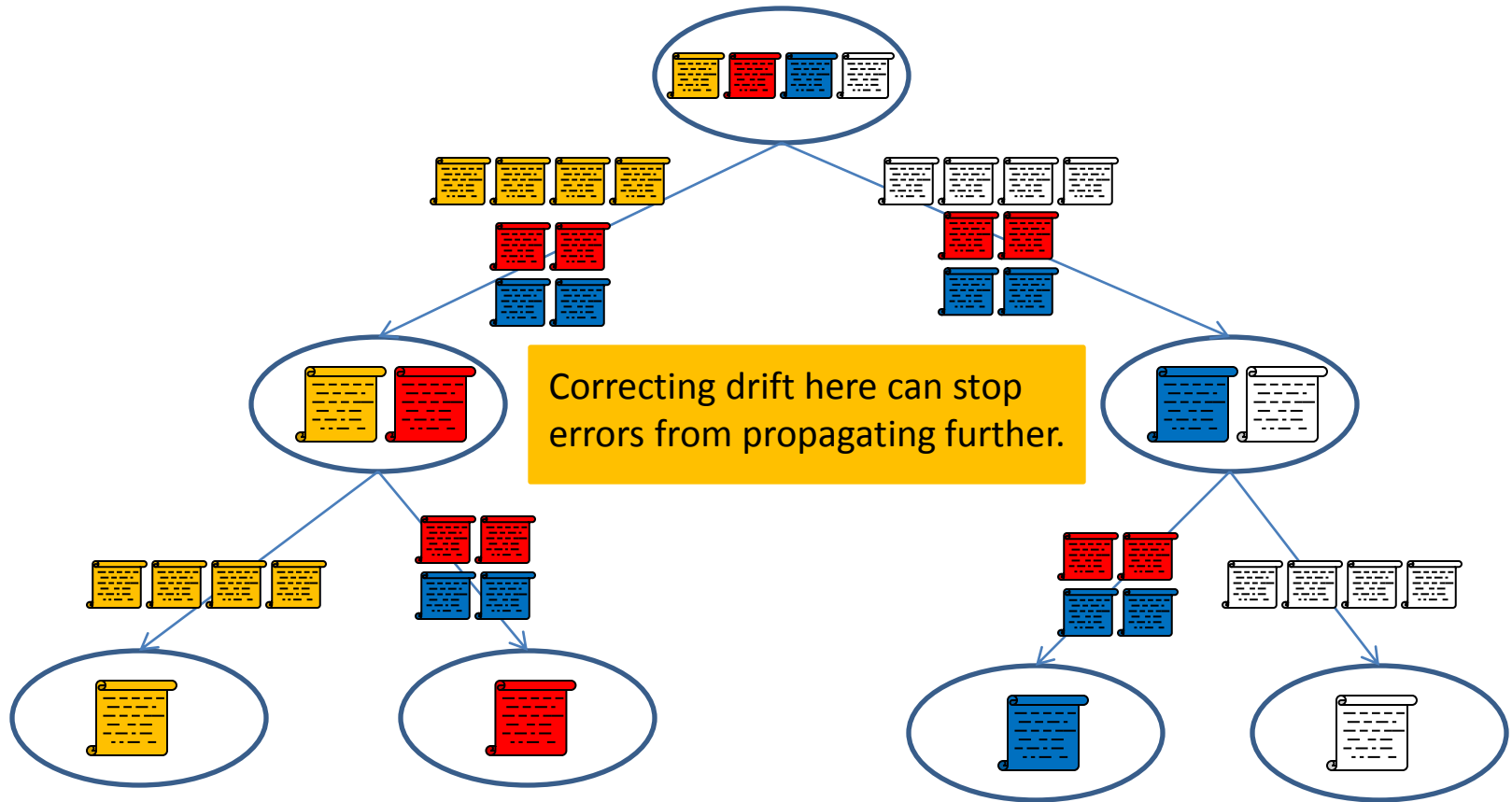
# Test Distribution Drift



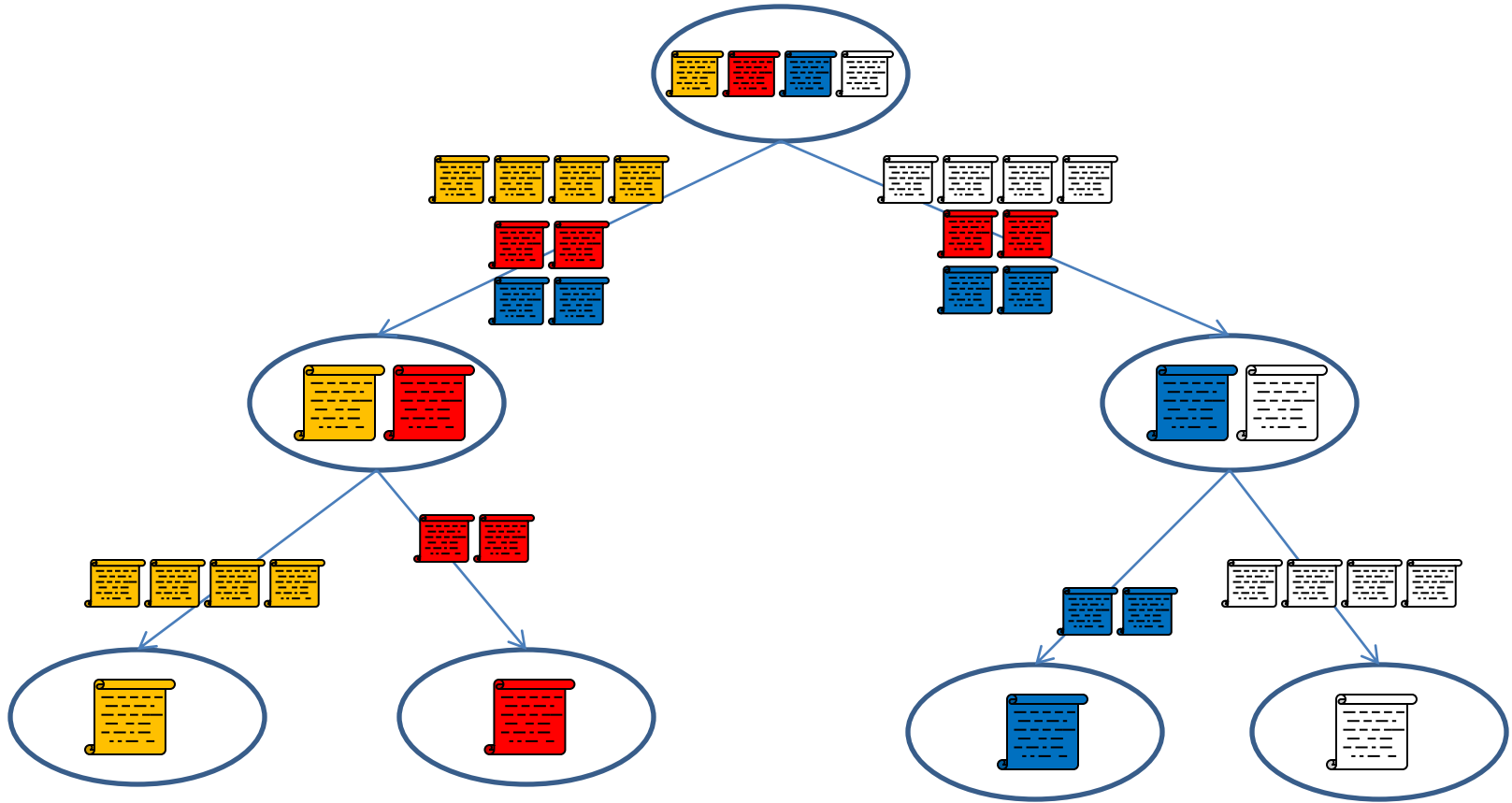
# Test Distribution Drift



# Test Distribution Drift



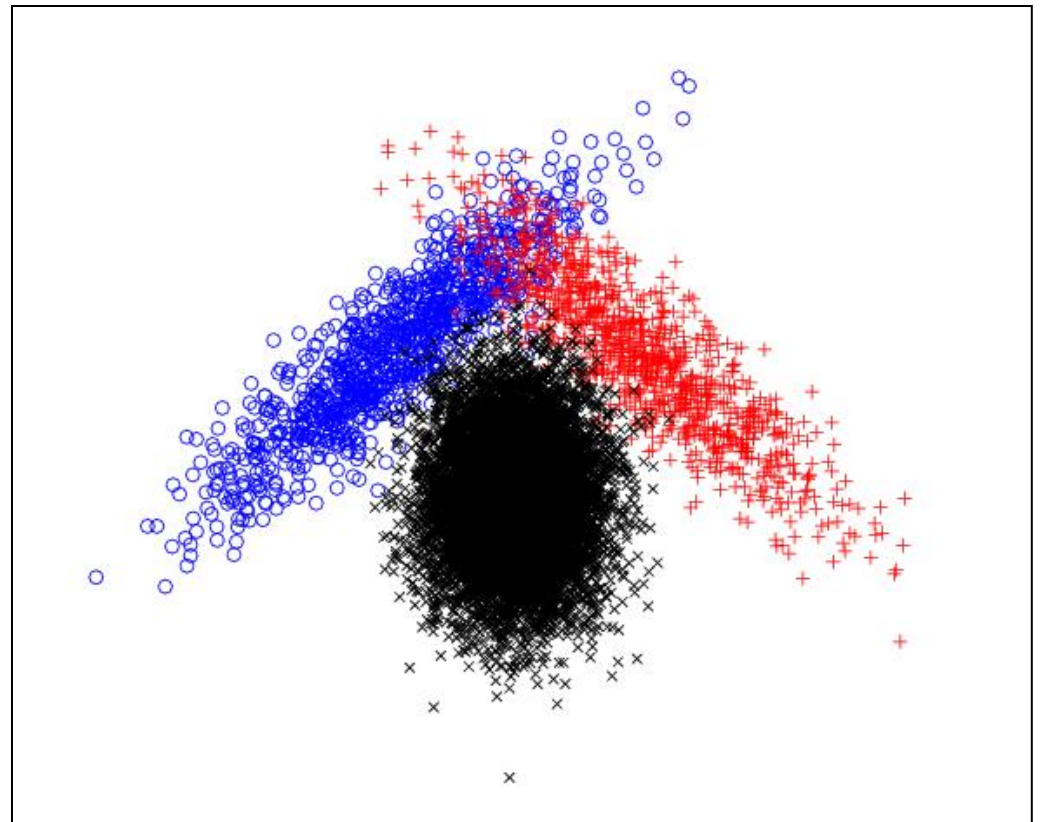
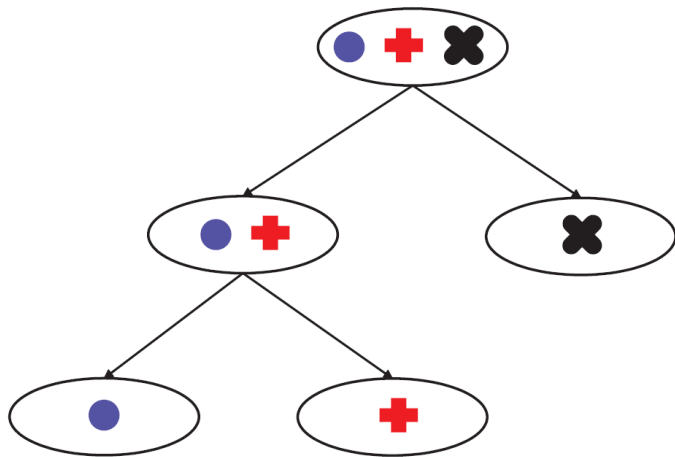
# Drift Corrected



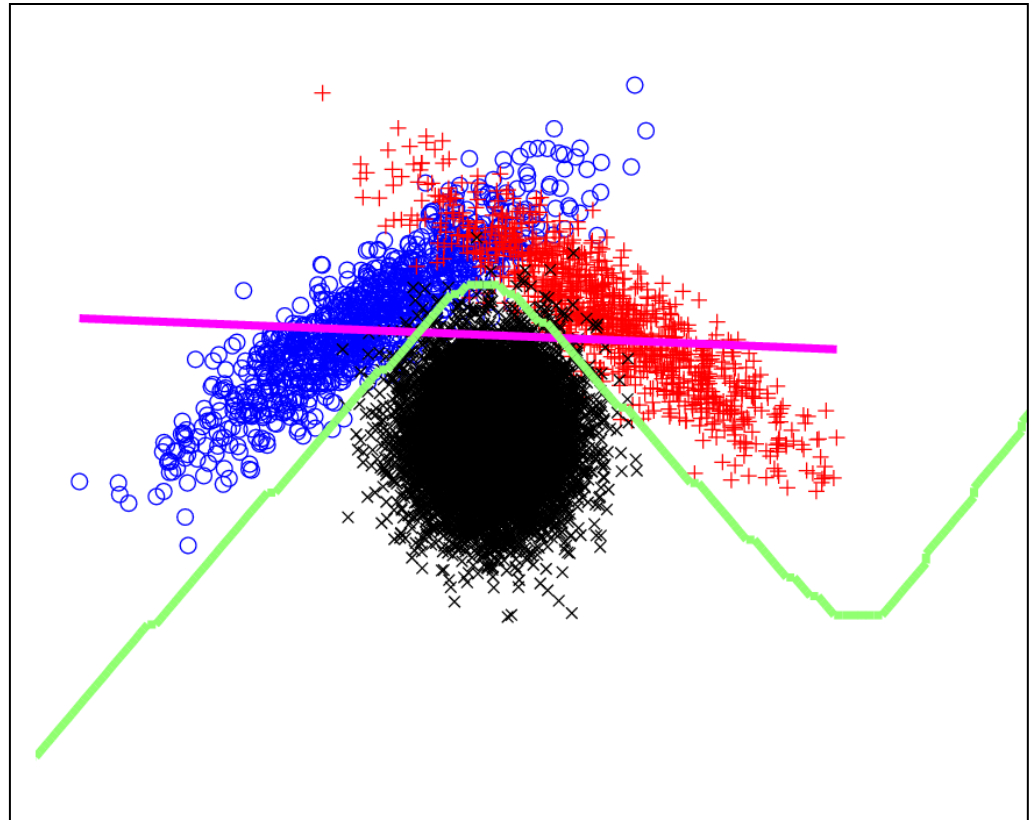
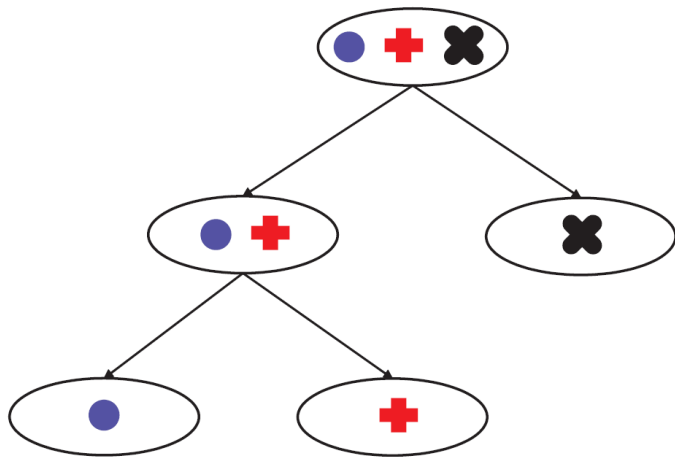
# Refinement

- Use cross-validation on the training set to predict test distribution.
- Training examples are *predicted* to be at node unioned with *actual* at node.
- Computational complexity
  - Training: Parallelization by level. Cost primarily cross-validation.
  - Testing: Prediction cost is the same.

# Complex Decision Surfaces



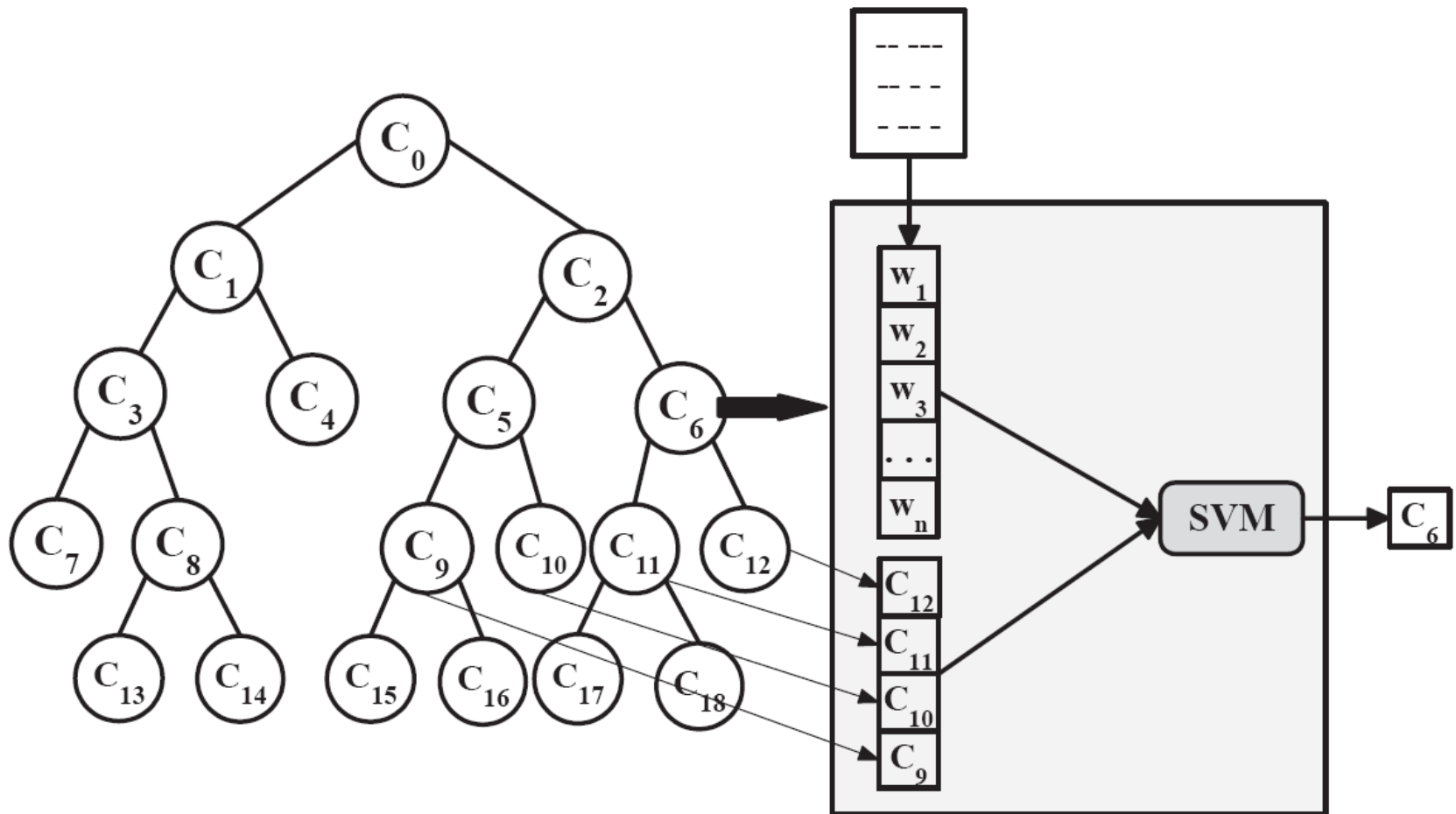
# Complex Decision Surfaces



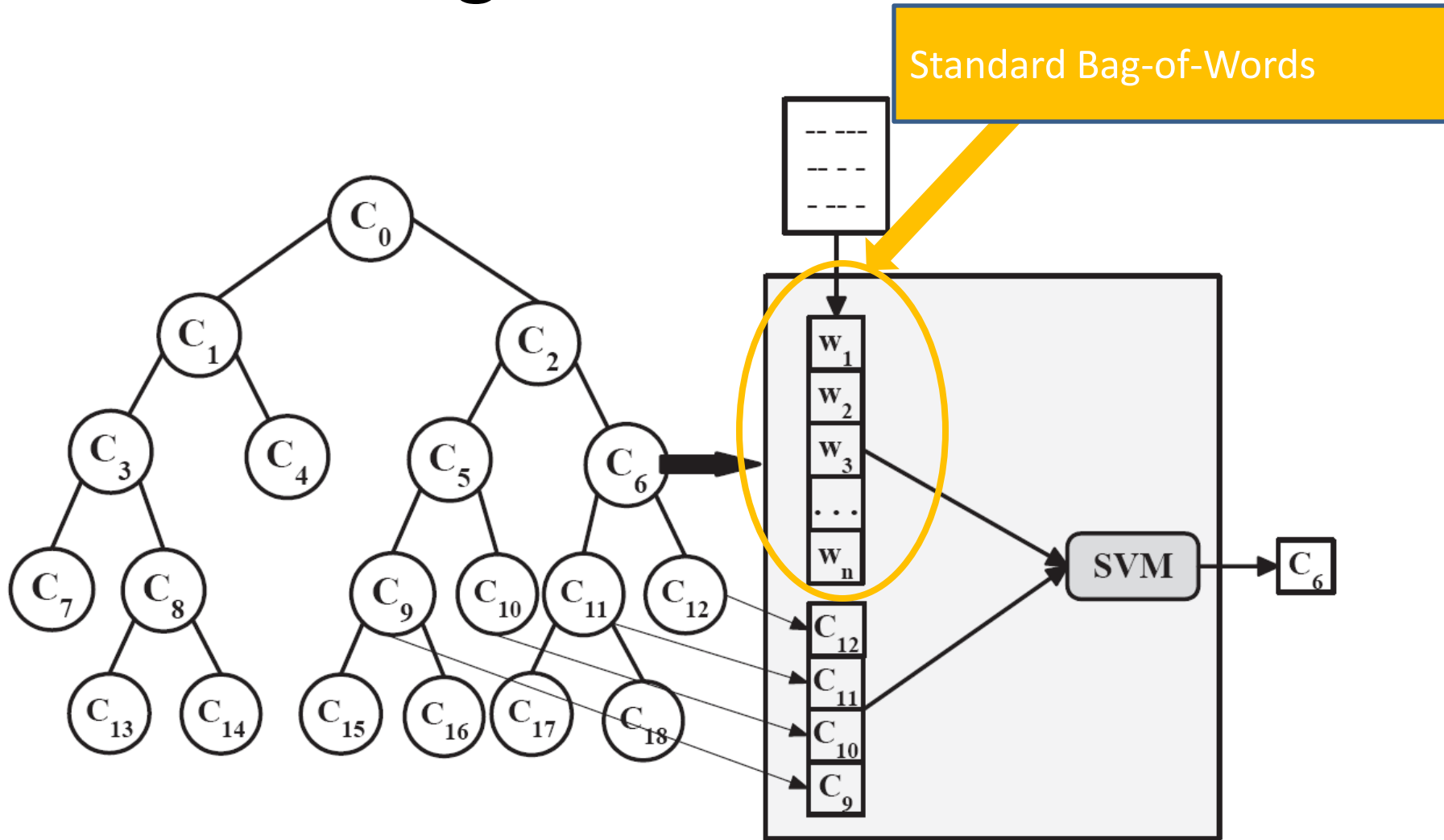
# Modeling Non-Linear Decision Surfaces

- Union of two or more linear surfaces is non-linear in general.
- Given a binary feature capturing whether a document belongs to a child, the surface becomes linear again.
- Introduce metafeatures where membership is predicted from the bottom-up before making a top-down final decision.

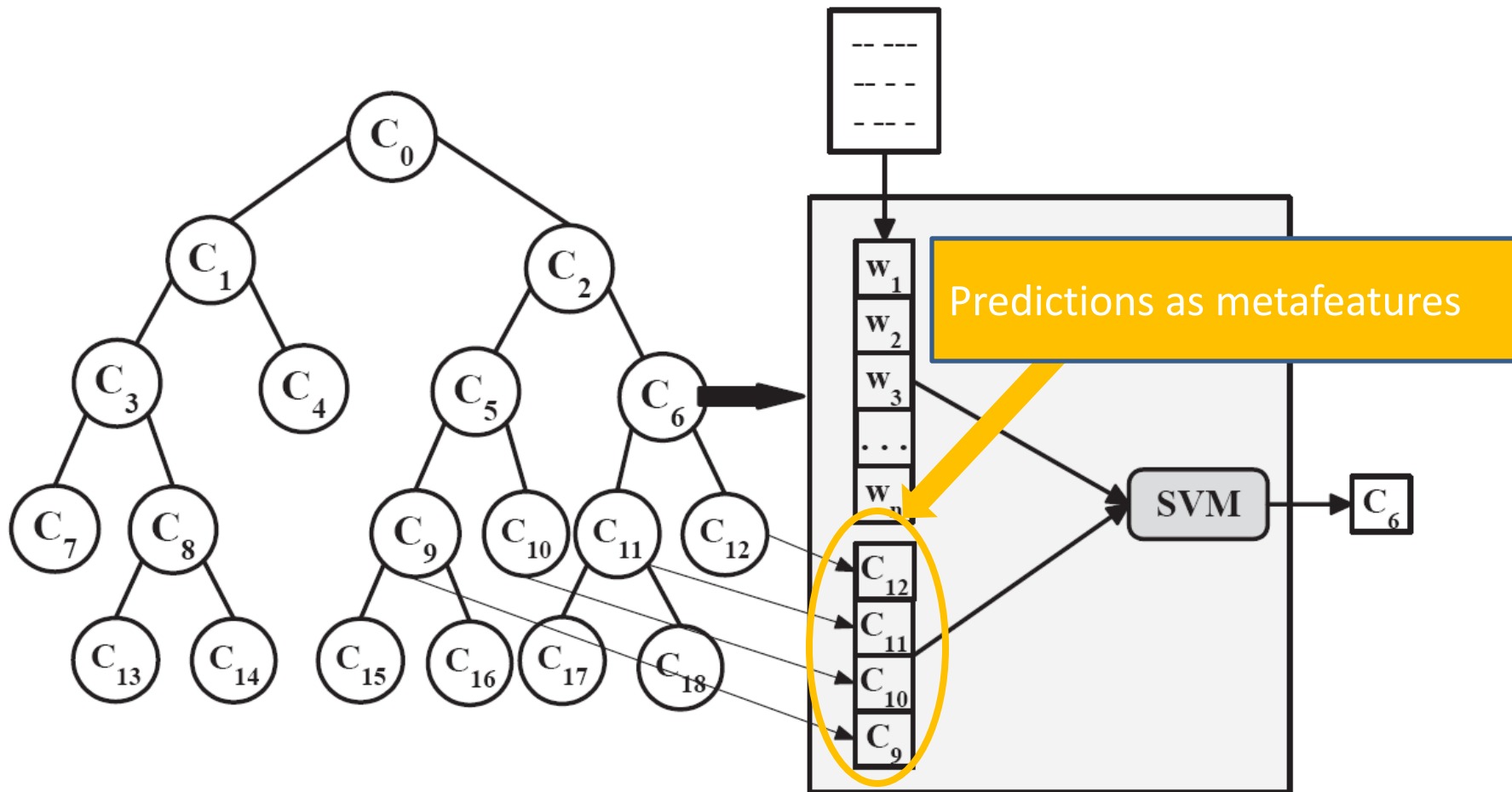
# Extending with Metafeatures



# Extending with Metafeatures



# Extending with Metafeatures



# Refined Experts

- Starting at leaves, use cross-validation to produce metafeatures for a node.
- At top, run refinement training algorithm augmented with metafeatures.
- Computational complexity
  - Training: Parallelization by level. Cost about 2x refinement.
  - Testing: Prediction cost can be much higher to “poll” experts from bottom. Can be distributed.

# Our Solution

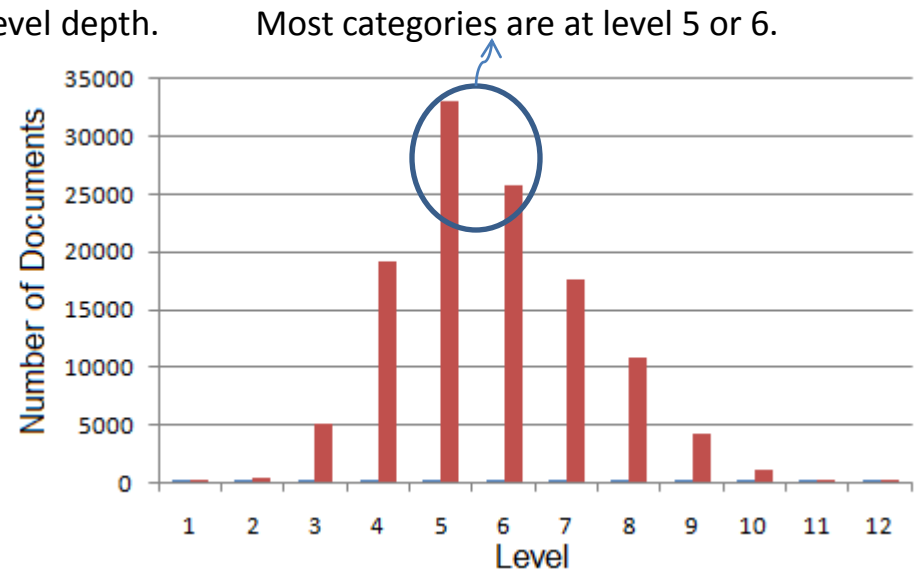
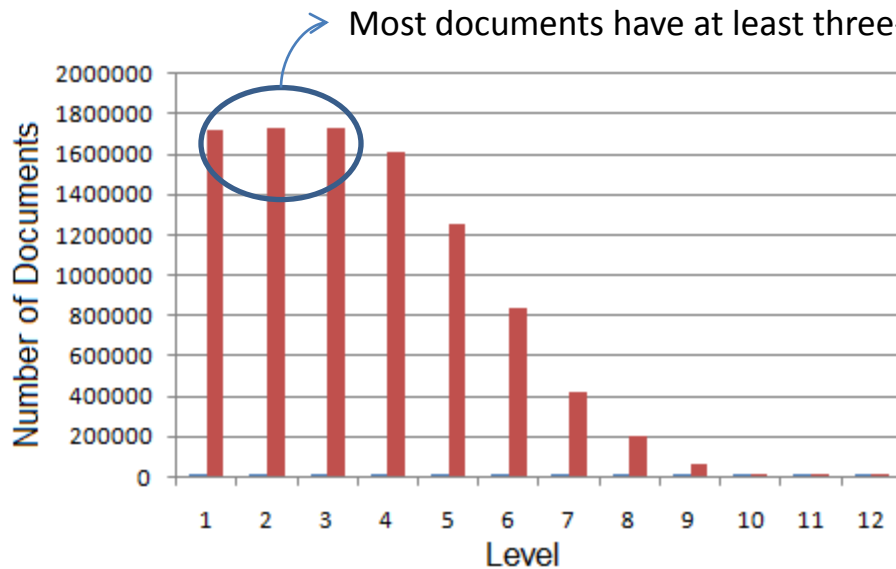
- Error Propagation
  - Predict the test time distribution at a node.
  - Incorporate when training model.
- Complex decision surfaces high in the hierarchy
  - Train “first guess” models as experts from bottom-up.
  - Make use of expert information when making final prediction top-down.

# Outline

1. Overview of Approach
2. Previous Work
3. Refinement & Refined Experts
4. Experiments and Results

# Data Set

- Data set: The Open Directory Project (ODP)
  - Number of Categories: 62,767
  - Number of Documents: 1,188,524/ 509,454 (train/test)



# Experiment Setup

- Base classifier: Linear-SVM
- Data representation: bag-of-words, tf-idf, unit sphere normalization.
- All methods optimize F1 (*Scut*)
- Selected 9 sub-hierarchies of ODP data
  - Adult, Computer, Game, Health, Home, Kids & Teens, Reference, Science, Shopping
  - 22635 categories total
- Here we present the average results for these sub-hierarchies.

# Halting Error Propagation

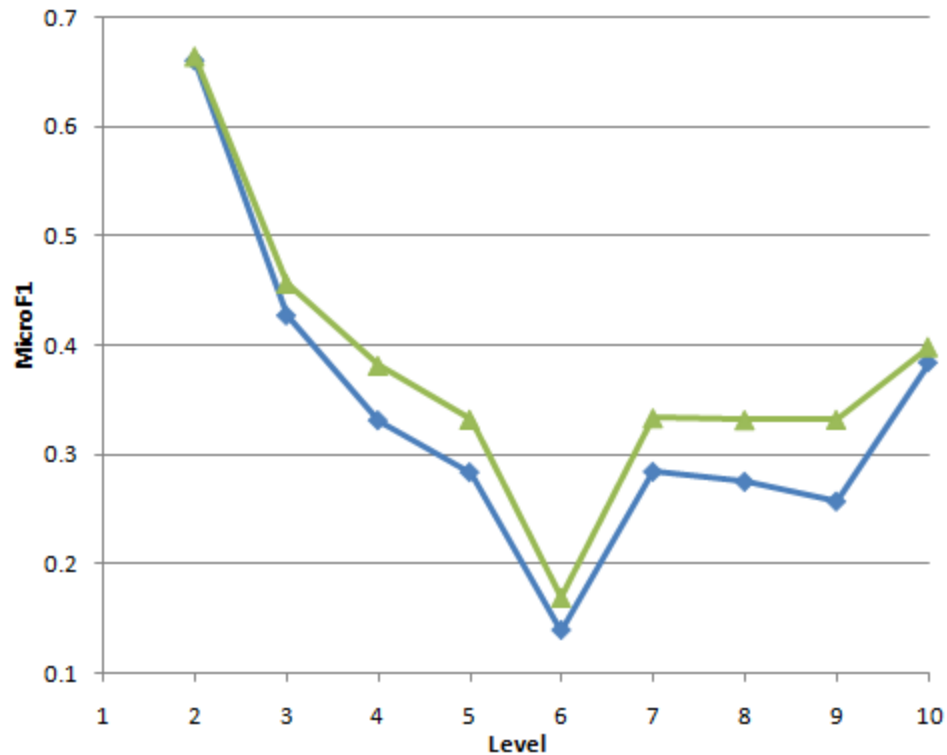
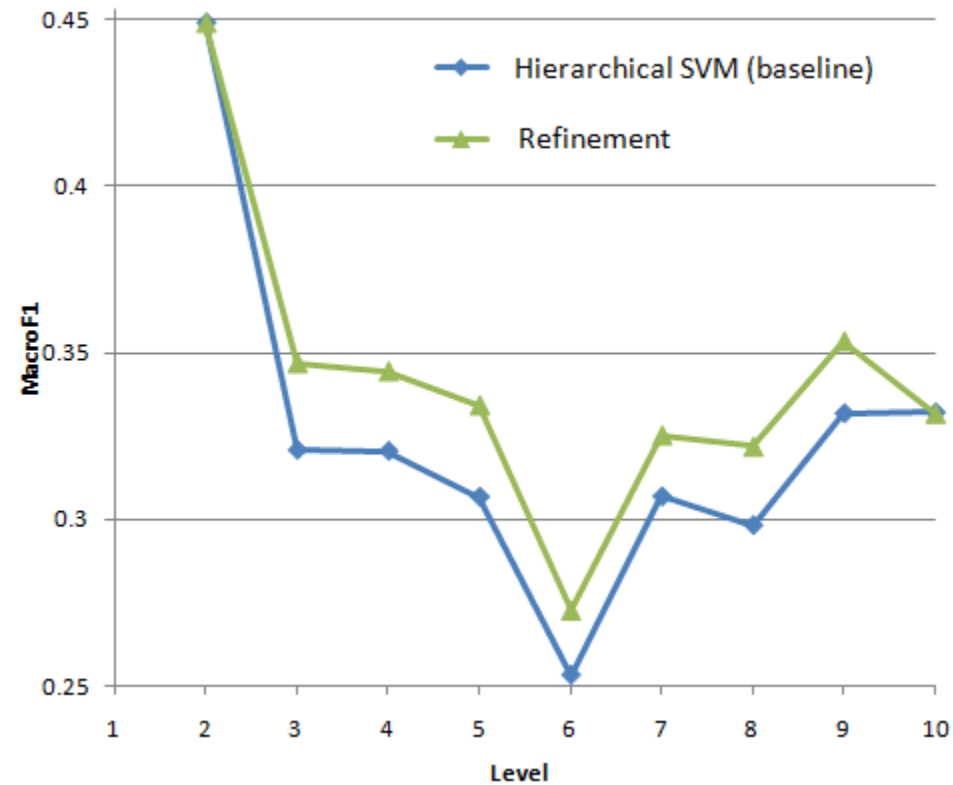
	Hierarchical SVM	Refinement
Macro F1	0.302	0.326 (7.9%)
Micro F1	0.365	0.414 (13.2%)

# Halting Error Propagation

	Hierarchical SVM	Refinement
Macro F1	0.302	0.326 (7.9%)
Micro F1	0.365	0.414 (13.2%)

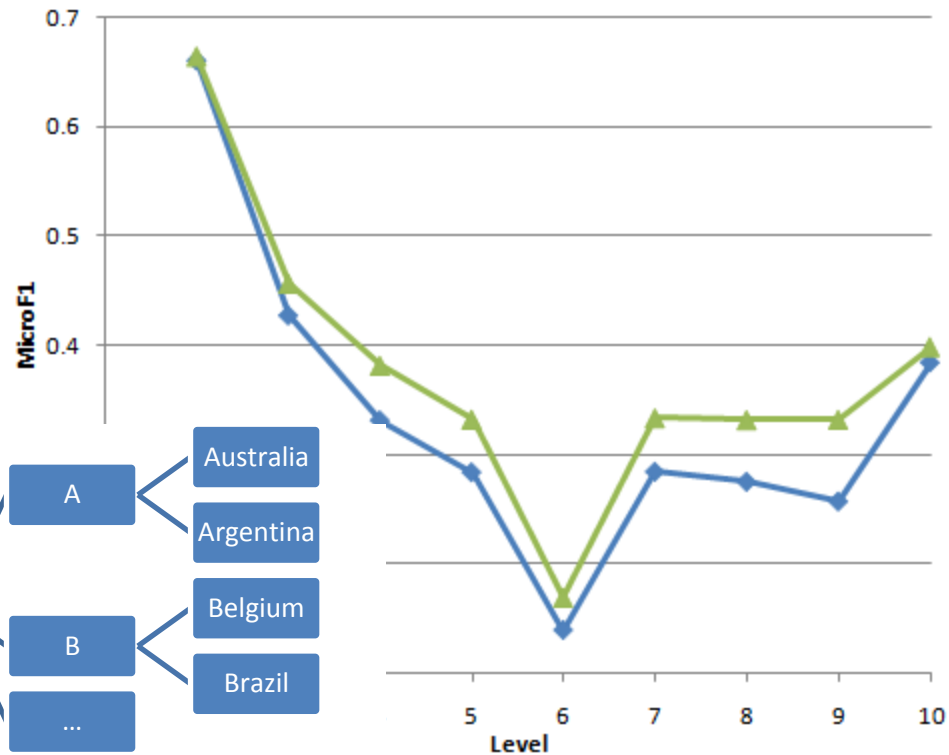
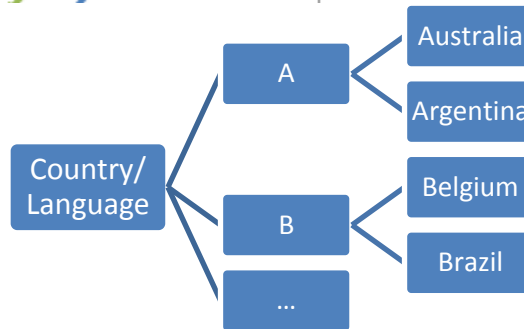
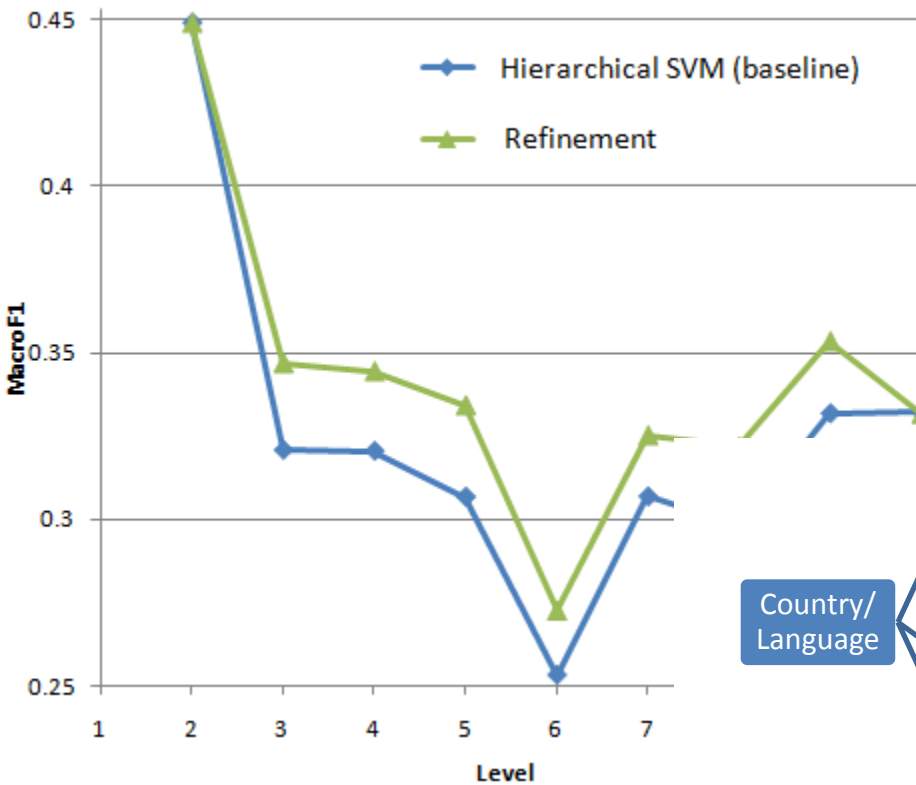
# Halting Error Propagation

	Hierarchical SVM	Refinement
Macro F1	0.302	0.326 (7.9%)
Micro F1	0.365	0.414 (13.2%)



# Halting Error Propagation

	Hierarchical SVM	Refinement
Macro F1	0.302	0.326 (7.9%)
Micro F1	0.365	0.414 (13.2%)



# Adding Expert Information

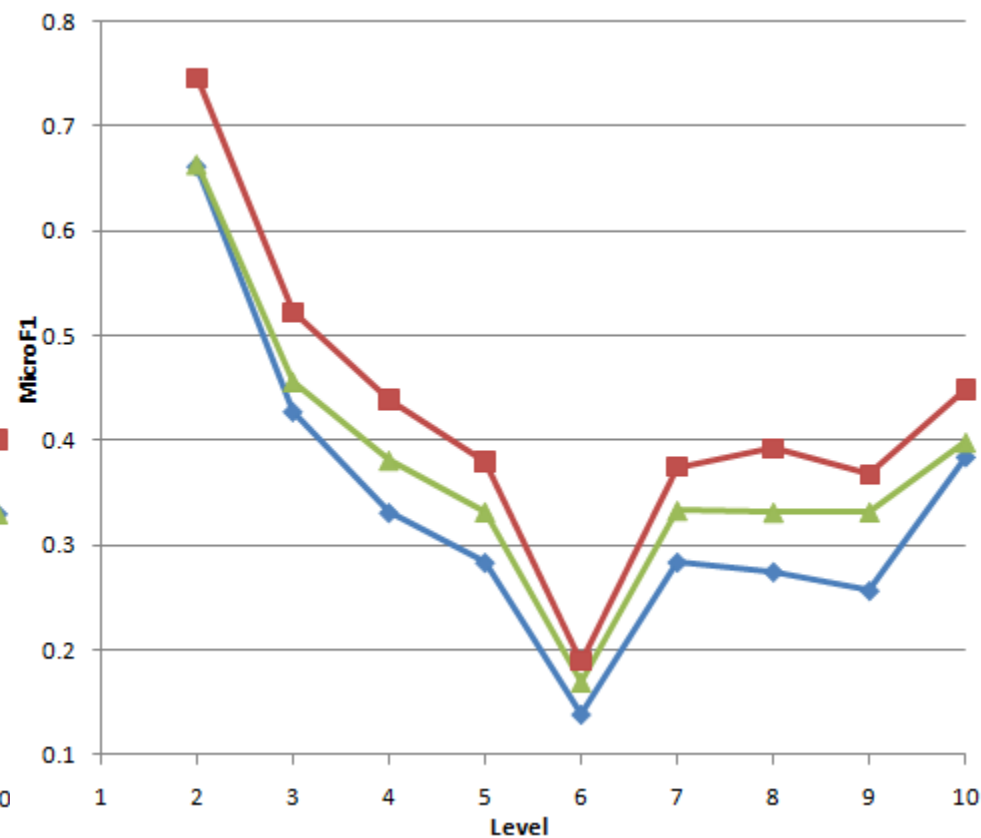
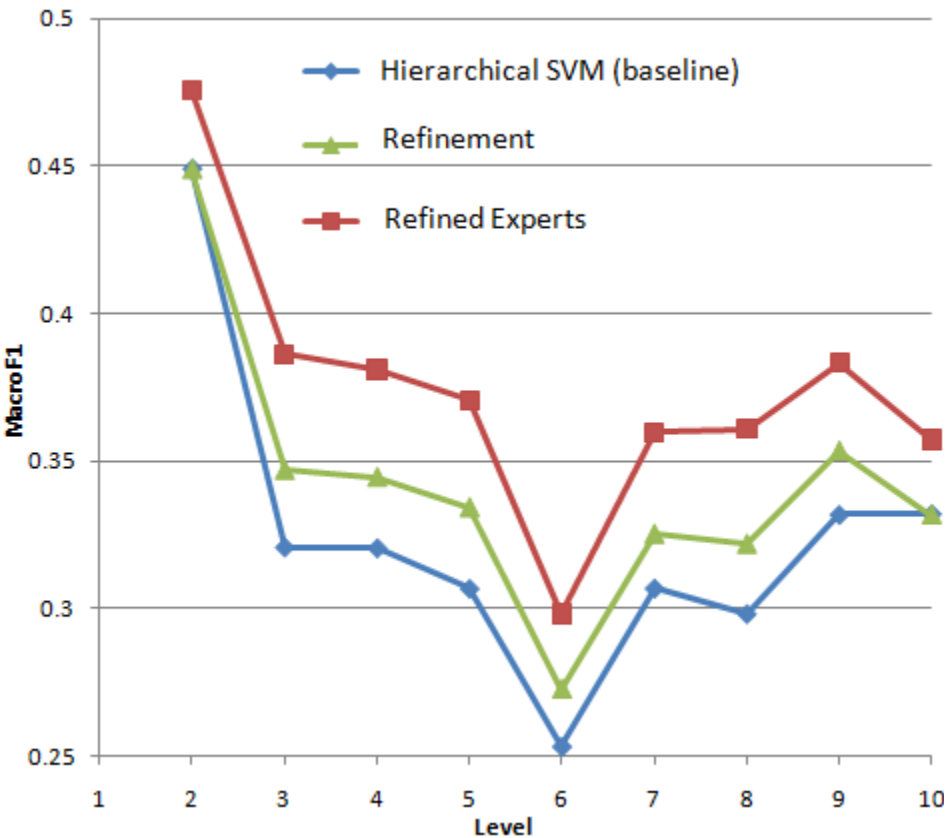
	Hierarchical SVM	Refinement	Refined Experts
Macro F1	0.302	0.326 (7.9%)	<b>0.361 (19.6%)</b>
Micro F1	0.365	0.414 (13.2%)	<b>0.468 (28.0%)</b>

# Adding Expert Information

	Hierarchical SVM	Refinement	Refined Experts
Macro F1	0.302	0.326 (7.9%)	<b>0.361 (19.6%)</b>
Micro F1	0.365	0.414 (13.2%)	<b>0.468 (28.0%)</b>

# Adding Expert Information

	Hierarchical SVM	Refinement	Refined Experts
Macro F1	0.302	0.326 (7.9%)	<b>0.361 (19.6%)</b>
Micro F1	0.365	0.414 (13.2%)	<b>0.468 (28.0%)</b>



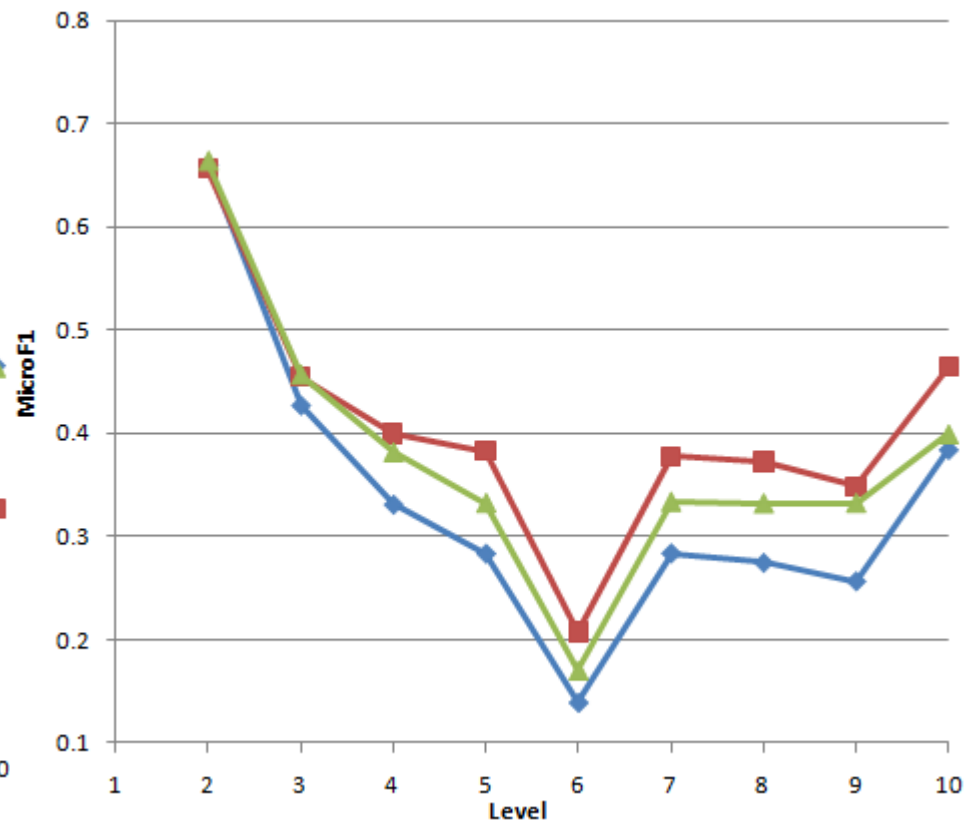
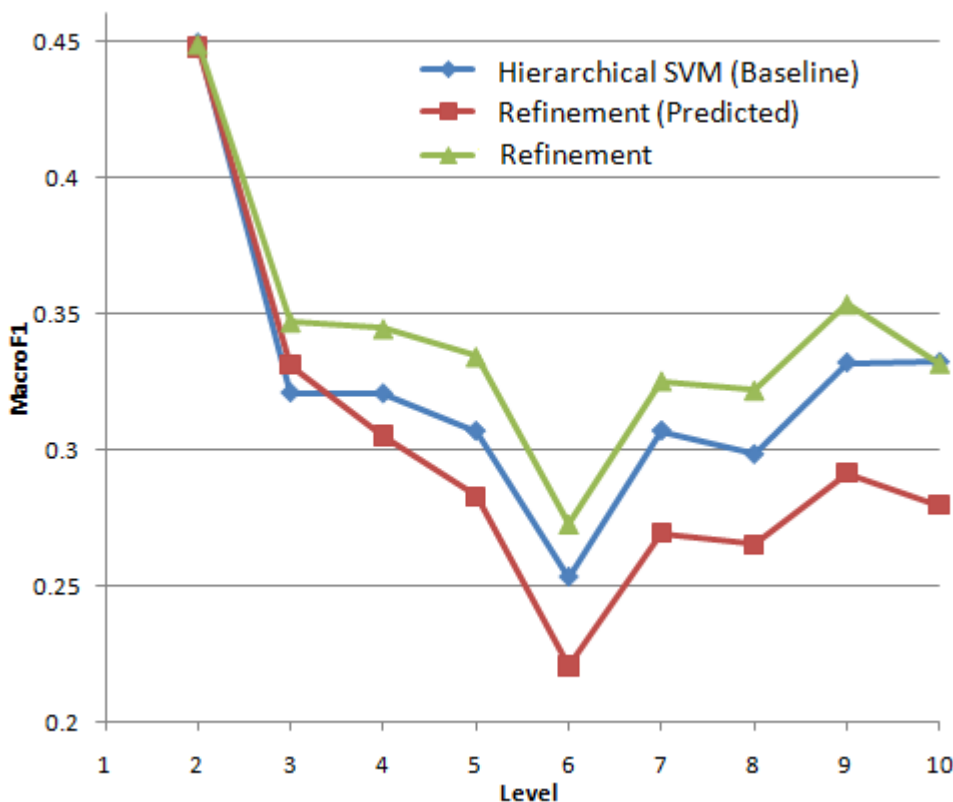
# Ablation Studies

- Use predicted distribution only at a node or predicted + actual?
- How much bottom-up information is useful?
- What type of expert information?
- Is non-linear transform of predictions useful?

# Ablation Studies

- Use predicted distribution only at a node or predicted + actual?
- How much bottom-up information is useful?
- What type of expert information?
- Is non-linear transform of predictions useful?

# Comparing Different Refinement Training Distributions



1. Training at Predicted Parent can improve microF1 but hurt macroF1 since there are a lot rare categories in the hierarchy.
2. Training with union improves in both microF1 and macroF1.

# Conclusion

- Refinement: addresses error propagation
  - **8-13% improvements overall**
  - Training is slightly more time
  - Prediction is computationally as efficient
- Refined Experts: Refinement + complex surfaces
  - **20-28% improvements overall**
  - More intensive training and testing
  - Highly parallelizable
- Future Work
  - Value of information by considering weights of bottom-up categories.
  - Use of insights in approximate graphical models.
- Our learning model is flexible. The base-classifier can be replaced by any other learners.

Thank you!

Questions?