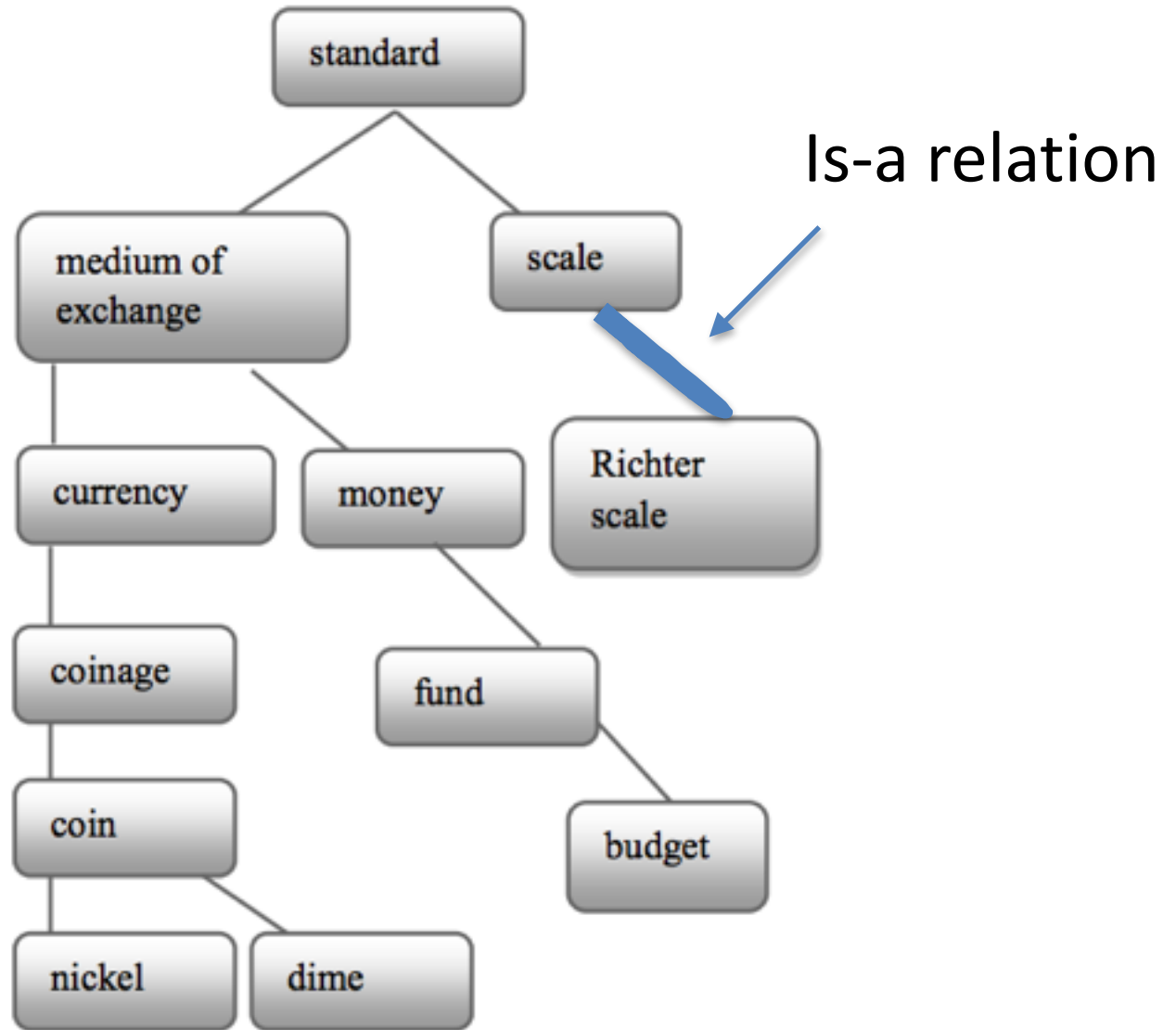


# similarity

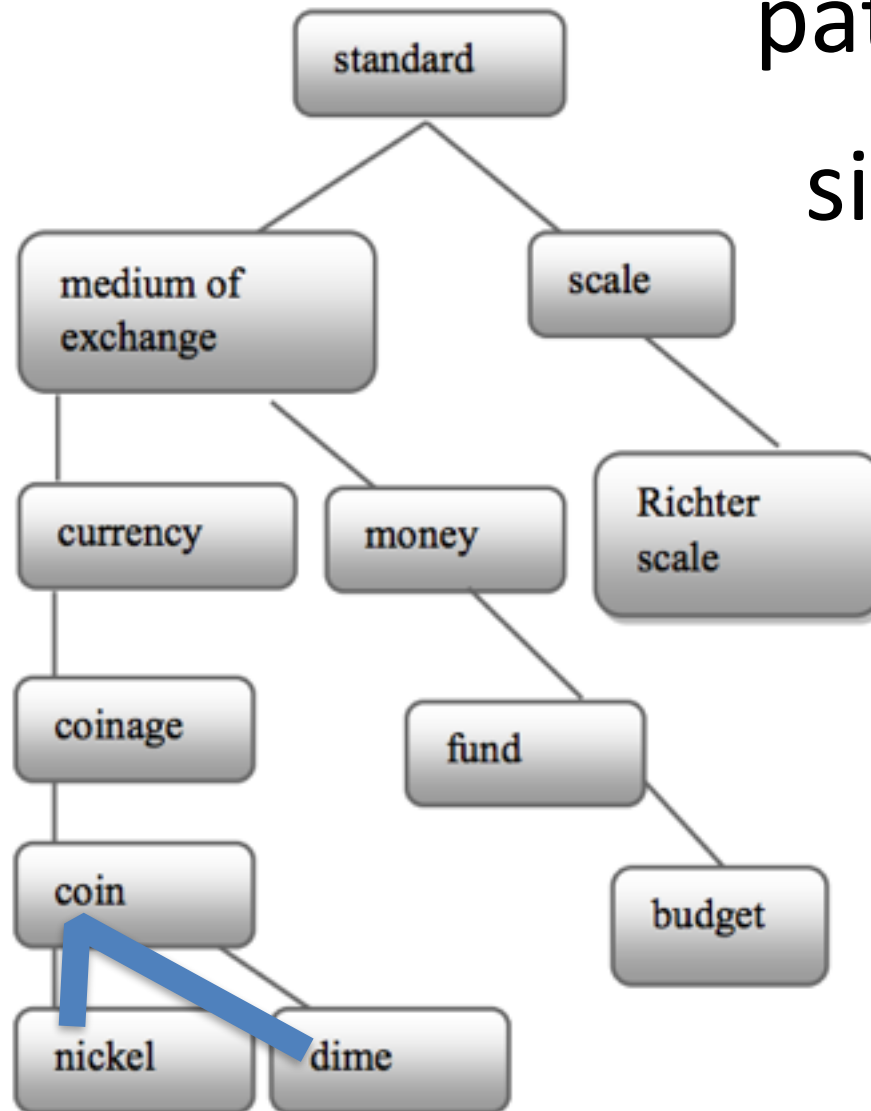


**Fig1: A fragment of the WordNet hypernym hierarchy**

Graph was taken from: Slimani, Thabet. "Description and evaluation of semantic similarity measures approaches." arXiv preprint arXiv: 1310.8059 (2013).

similarity  
path

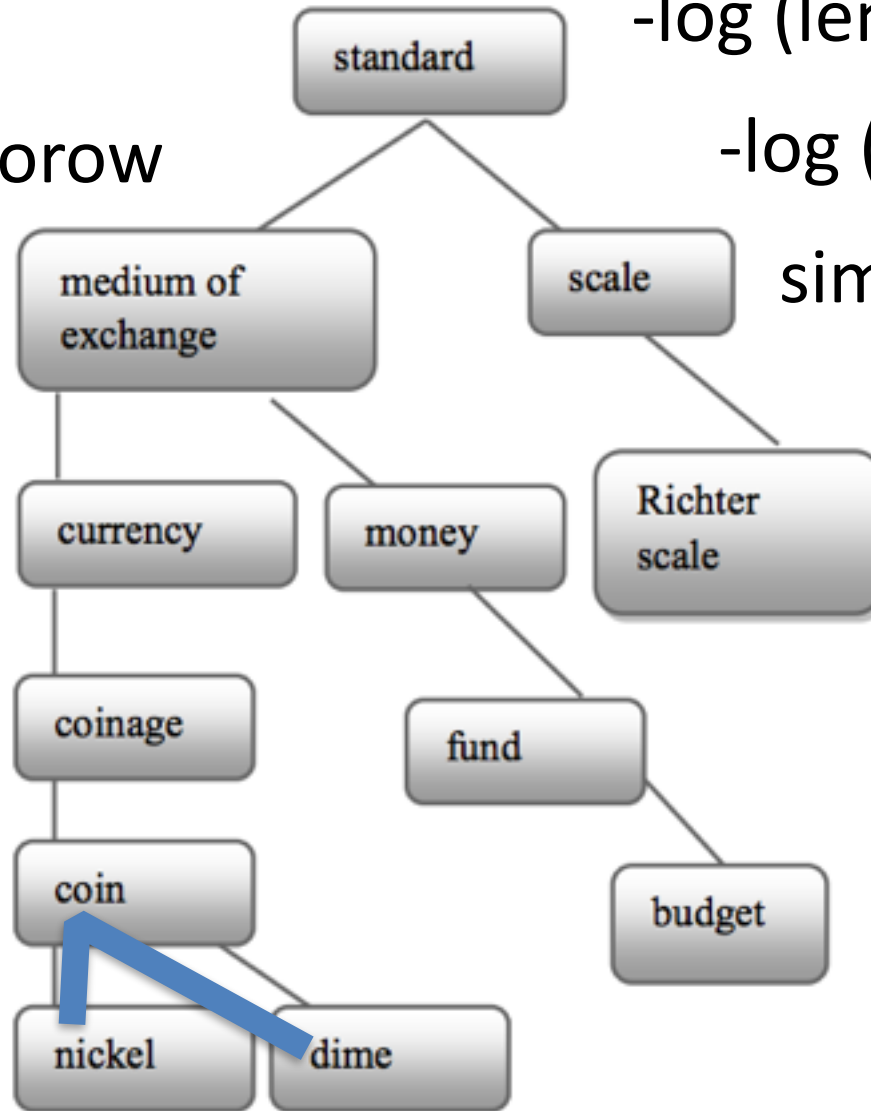
path: 3 steps  
 $\text{sim} = 1/3$



**Fig1: A fragment of the WordNet hypernym hierarchy**

# similarity

Leacock & Chodorow



$$-\log (\text{length} / (2 * D))$$

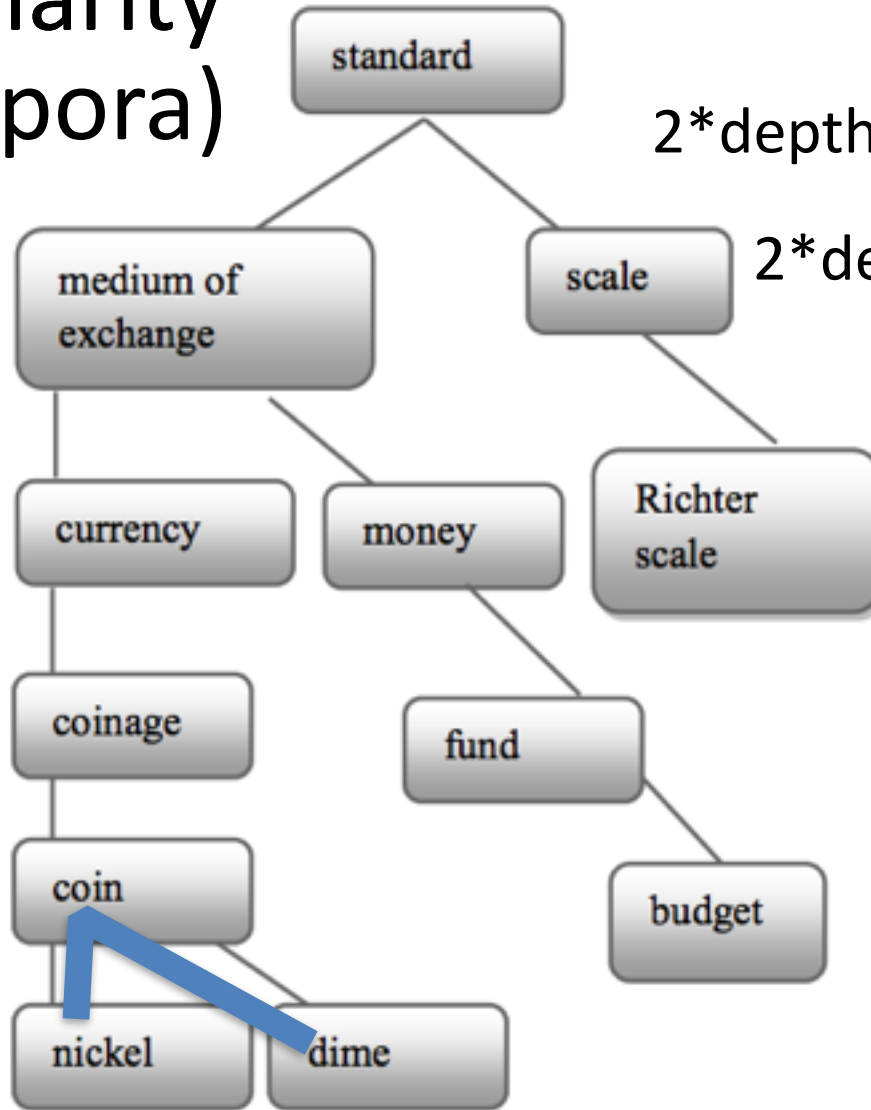
$$-\log (3 / (2 * 6))$$

$$\text{sim} = 1.39$$

**Fig1: A fragment of the WordNet hypernym hierarchy**

similarity  
(corpora)

Wu & Palmer



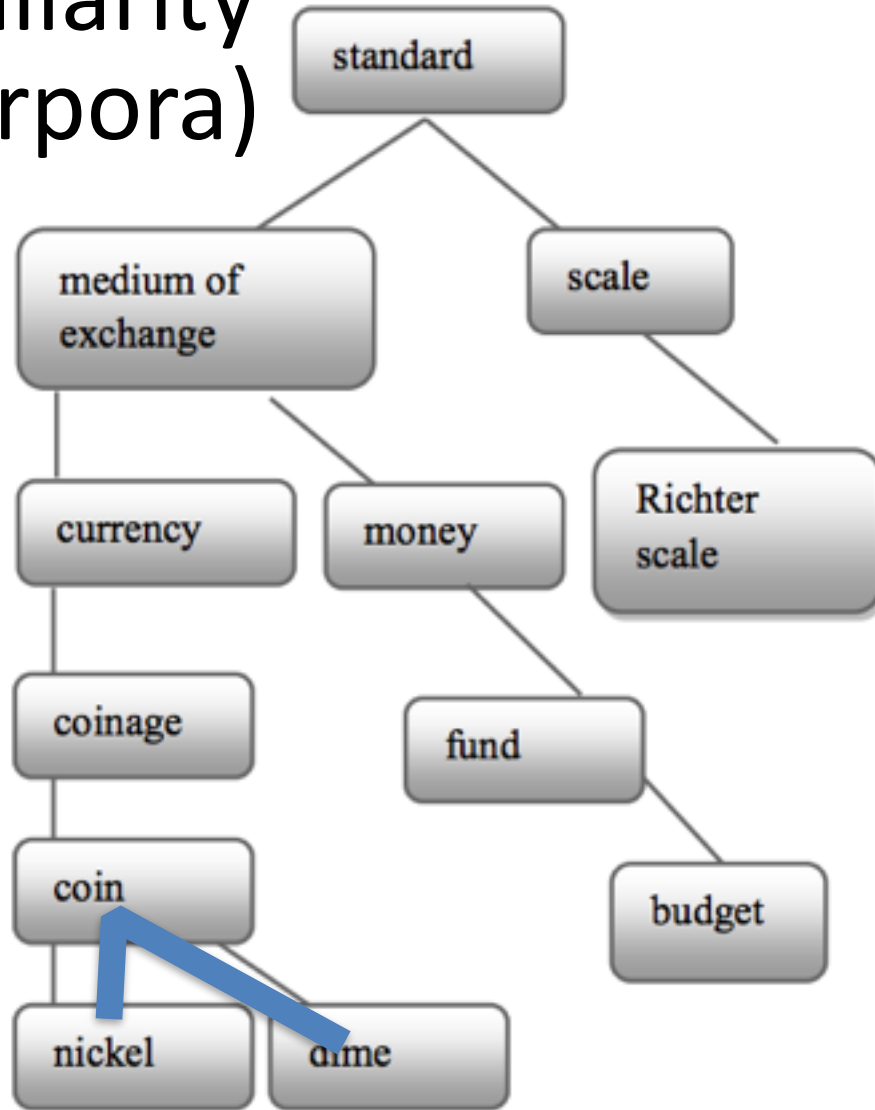
$$2 * \text{depth}(\text{lcs}) / (\text{depth}(s1) + \text{depth}(s2))$$

$$2 * \text{depth}(5) / (\text{depth}(6) + \text{depth}(6))$$

$$\text{Sim} = 0.83$$

**Fig1: A fragment of the WordNet hypernym hierarchy**

similarity  
(corpora)



Resnik

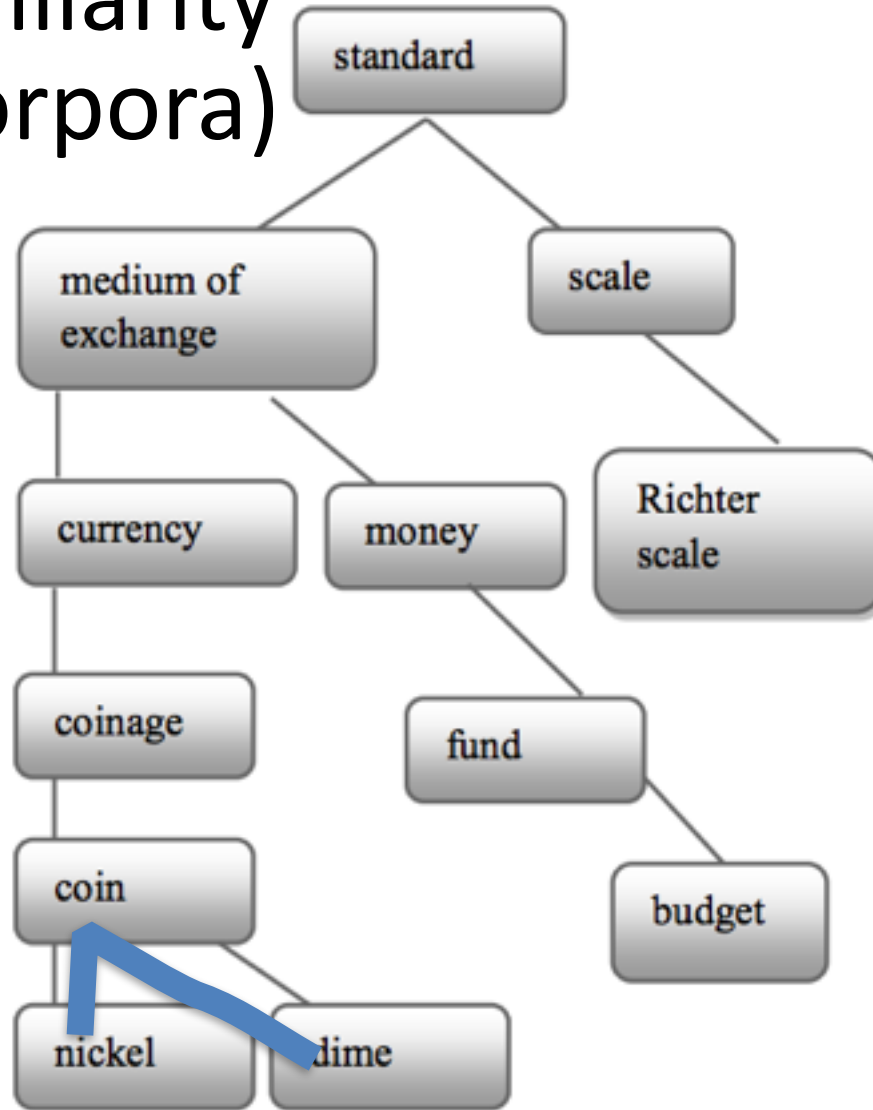
$\text{sim} = \text{IC}(\text{lcs})$

$\text{IC} = -\log(p)$

$\text{sim} = -\log(0.002)$   
 $= 6.21$

**Fig1: A fragment of the WordNet hypernym hierarchy**

similarity  
(corpora)



Lin

Jiang & Conrath

also use prob of  
senses:  
(nickel and dime)

**Fig1: A fragment of the WordNet hypernym hierarchy**

relatedness  
(definitions)

Lesk  
overlap in words

a United States coin worth one tenth  
of a dollar

a hard malleable ductile silvery  
metallic element that is resistant to  
corrosion; used in alloys; occurs in  
pentlandite and smaltite and garnierite  
and millerite

many variations using all  
related senses around words

Similarity: is-a relation

Relatedness: all semantic relations  
or definitions

use external corpora (state of the art)



# Main idea: state of art uses ideas from “old” algorithms

An Enhanced Lesk Word Sense Disambiguation Algorithm through a Distributional Semantic Model (state of the art on one dataset)

Unsupervised Most Frequent Sense Detection using Word Embeddings