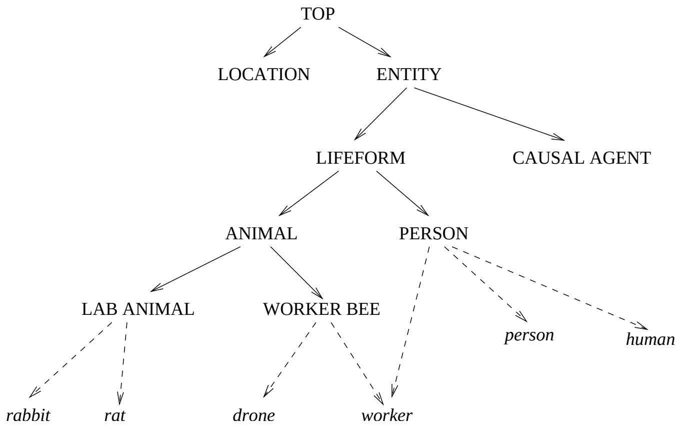
CS674, Spring 2002
Lecture Handout, 3/6/02

1. OED definitions of two different senses of "bank".

1**(a)** The shelving or sloping margin of a river or stream; the ground bordering upon a river. [Or,] A raised or rising edge or margin of a pond, lake, pit, road, railway cutting, or other hollow place; in Mining, the surface of the ground at the pit-mouth, or top of the shaft.

1**(b)** An establishment for the custody of money received from, or on behalf of, its customers. Its essential duty is the payment of the orders given on it by the customers; its profits arise mainly from the investment of the money left unused by them.

2. Sample portion of WordNet (adapted from Abney and Light 1999).



3. The (Shannon/average) mutual information between two discrete random variables X and Y taking values from $\mathcal{X}$ and $\mathcal{Y}$, respectively:

$$\begin{aligned} I(X,Y) &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(X=x,Y=y)\log\frac{P(X=x,Y=y)}{P(X=x)P(Y=y)} \\ &= E_{p(X,Y)}\left[\log\frac{P(X=x,Y=y)}{P(X=x)P(Y=y)}\right] \end{aligned}$$

Compare the pointwise/Fano mutual information between two *values* x and y:

$$\hat{I}(x,y) = \log\frac{P(X=x,Y=y)}{P(X=x)P(Y=y)}.$$

4. The log-likelihood ratio for sense s_i and feature value x_j (e.g. Yarowsky (1995)):

$$\log\frac{P(s_i|x_j)}{P(\neg s_i|x_j)}$$

(Note: the *log-likelihood ratio test* is usually defined as testing a hypothesis regarding the values of parameters of a generative model of the data.)