Quiz #1:

1 point

1. Which of the following types of ambiguity were *not* discussed in the lecture?

- part of speech ambiguity
- syntactic ambiguity
- statistical ambiguity
- morphological ambiguity
- sense ambiguity

1 point

2. The sentence "My aunt collects old clocks and silverware" illustrates what type of ambiguity?

- Coordinating conjunction (CC) ambiguity
- Dialogue ambiguity
- Adjective phrase (AP) ambiguity
- Prepositional phrase (PP) ambiguity
- The sentence is not ambiguous.

1 point

3. Which of the following sentences involves a ditransitive verb?

- Who found the key?
- The dog sleeps.
- Mary gave the dog a bone.
- Wake up!
- The dog chased the cat.

1 point

4. Which of the following is *not* a noun phrase?

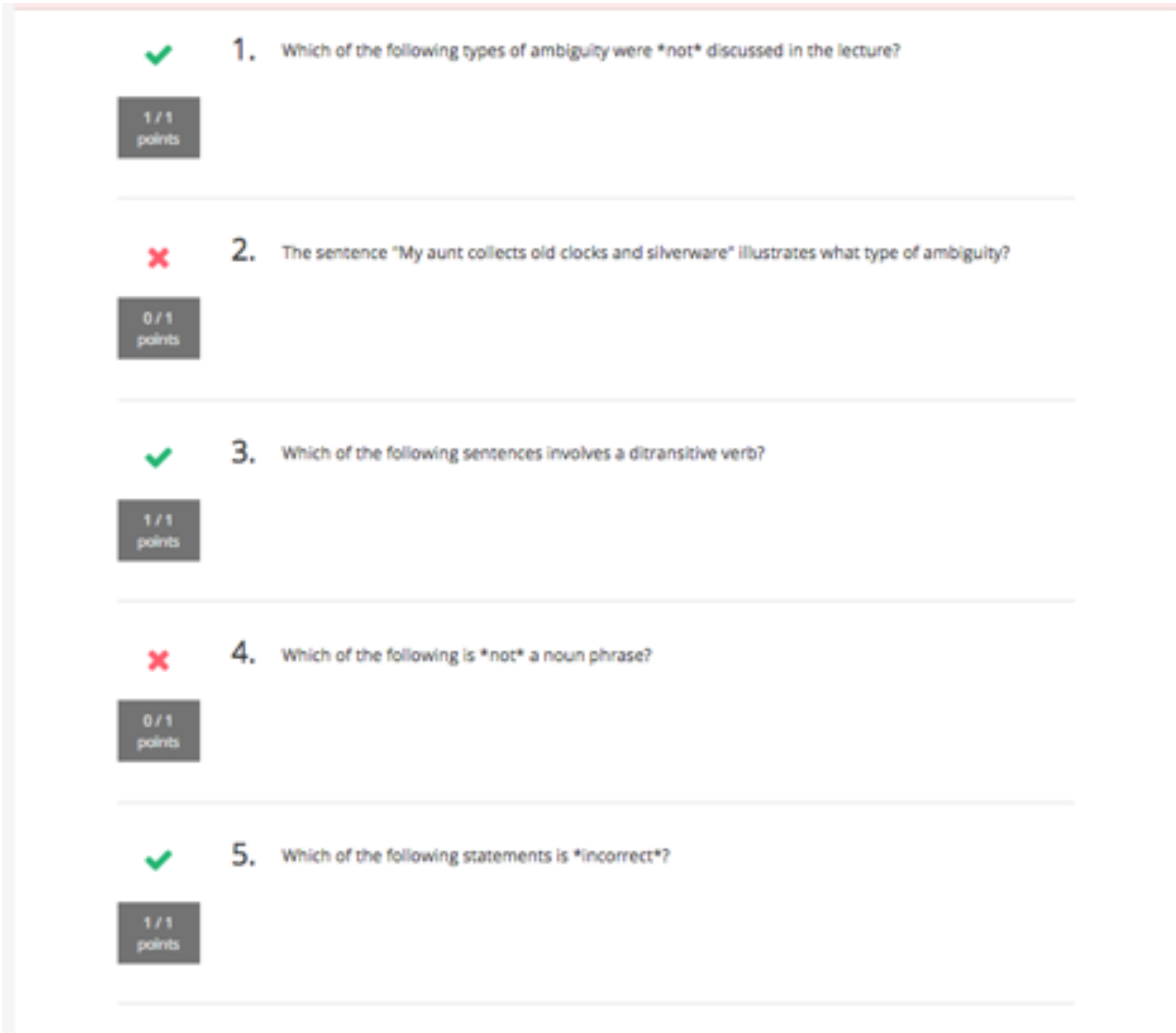
- ○ One of the most famous enterpreneurs in the world
- ○ People who speak Japanese
- ○ She
- ○ The children ate the oranges
- ◉ My father's friend

1 point

5. Which of the following statements is *incorrect*?

- ◉ Pragmatics is the study of the applications of finite-state automata to language
- ○ Morphology is the study of word components
- ○ Lexical semantics is the study of the meanings of words
- ○ Phonetics is the study of sounds
- ○ Syntax is the study of sentence and phrase structure

Graded Solution:



1. Which of the following types of ambiguity were *not* discussed in the lecture?

1 / 1 points

2. The sentence "My aunt collects old clocks and silverware" illustrates what type of ambiguity?

0 / 1 points

3. Which of the following sentences involves a ditransitive verb?

1 / 1 points

4. Which of the following is *not* a noun phrase?

0 / 1 points

5. Which of the following statements is *incorrect*?

1 / 1 points

#2: the sentenct my aunt collects old clocks and silverware:

what type of ambuiity: picked no ambiguity which was wrong then PP ambiguity on who is the aunt which was wrong. Next pick coordinating conjunction which is if the silveware is old or not. We know the clocks are old. If this is correct then this is a lame question.

#4: She picked and My fathers friend as not noun phrases. Both wrong.

Attempt #2:

1 point

1. Which of the following types of ambiguity were *not* discussed in the lecture?

- ( ) morphological ambiguity
- ( ) sense ambiguity
- ( ) part of speech ambiguity
- ( ) syntactic ambiguity
- (x) statistical ambiguity

1 point

2. The sentence "My aunt collects old clocks and silverware" illustrates what type of ambiguity?

- ( ) Adjective phrase (AP) ambiguity
- (x) Coordinating conjunction (CC) ambiguity
- ( ) The sentence is not ambiguous.
- ( ) Dialogue ambiguity
- ( ) Prepositional phrase (PP) ambiguity

1 point

3. Which of the following sentences involves a ditransitive verb?

- ( ) Who found the key?
- ( ) The dog chased the cat.
- ( ) The dog sleeps.
- (x) Mary gave the dog a bone.
- ( ) Wake up!

1 point

4. Which of the following is *not* a noun phrase?

- ( ) She
- ( ) One of the most famous enterpreneurs in the world
- ( ) The children ate the oranges
- (x) People who speak Japanese
- ( ) My father's friend

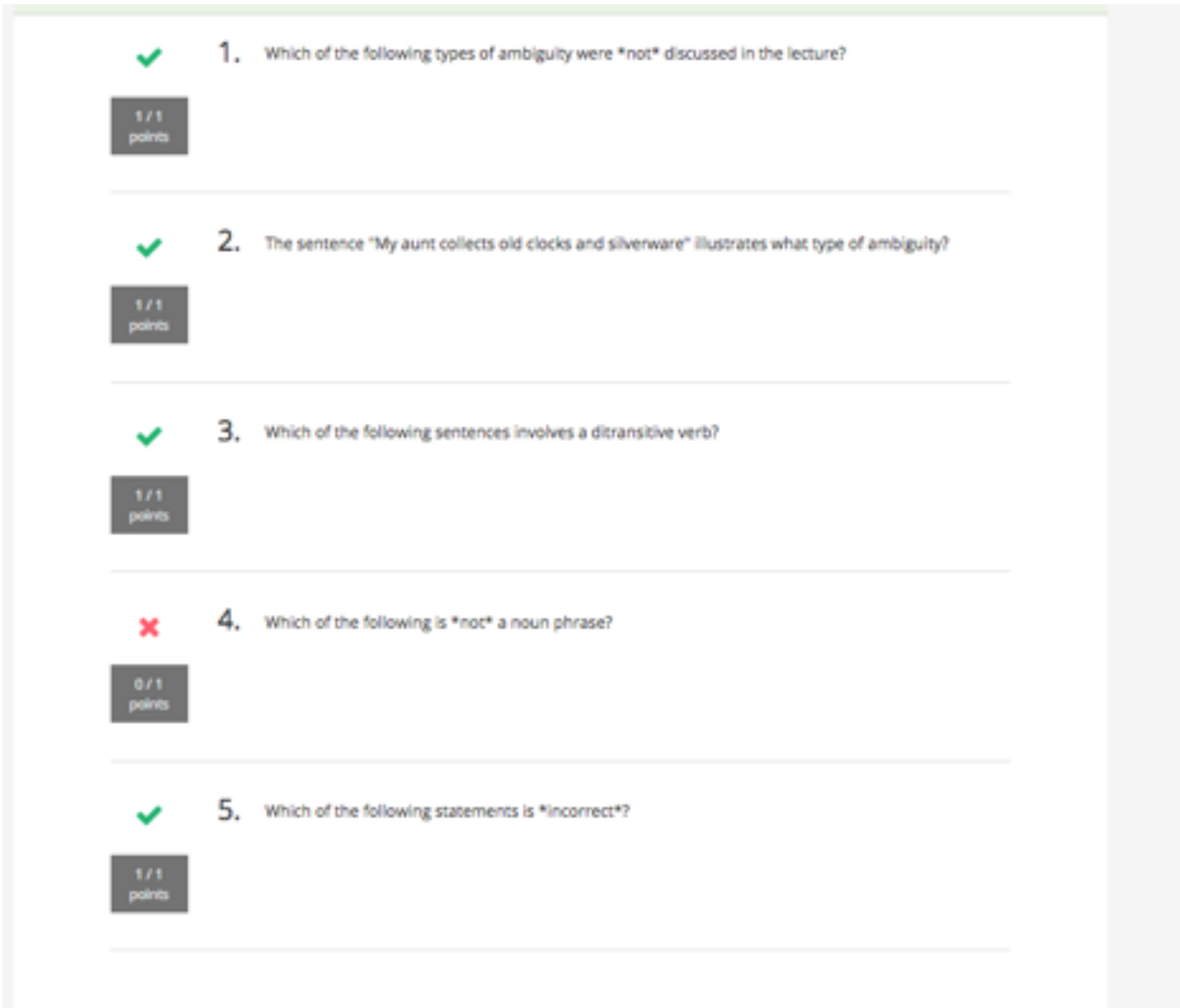
1 point

5. Which of the following statements is *incorrect*?

- ( ) Lexical semantics is the study of the meanings of words
- (x) Pragmatics is the study of the applications of finite-state automata to language
- ( ) Morphology is the study of word components
- ( ) Phonetics is the study of sounds
- ( ) Syntax is the study of sentence and phrase structure

OK for the not noun phrase we tried 1) she. was wrong. I assumed this was a PP; and wasnt a NP even though PP are considered NPs. 2) My father's friend; refers to an entity even there is no verb. this was wrong 3) People who speak japanese (plural, a group of people; not a single person or thing. Verb is speak. Can say Japanese speaking people. Is this a valid noun phrase?

REsults:



1. Which of the following types of ambiguity were *not* discussed in the lecture?
1 / 1 points

2. The sentence "My aunt collects old clocks and silverware" illustrates what type of ambiguity?
1 / 1 points

3. Which of the following sentences involves a ditransitive verb?
1 / 1 points

4. Which of the following is *not* a noun phrase?
0 / 1 points

5. Which of the following statements is *incorrect*?
1 / 1 points

4 is still wrong. Try one more time after time; the one of the most famous… this is really stupid.

Week 2 Quiz:

1 point

1. The dependency representation of the sentence "Jane has a cat" is the following (with the arrows pointing from parent to child node):

- cat -> has, Jane -> cat, a -> cat
- Jane -> has, has -> a, a -> cat
- has -> Jane, Jane -> cat, a -> cat
- has -> Jane, Jane -> cat, cat -> has
- has -> cat, has -> Jane, cat -> a

To perform dependency representation we start from the verb.
Verb is has so has-> Jane and has->cat. a is associated with cat. Last one is ans.

1 point

2. Which of the following statements is true:

- In Semantic Role Labeling, the verb "accept" takes exactly two arguments (A0=acceptor) and (A1=thing accepted).
- The word "deer" is a hypernym of the word "animal"
- The English verb "play" can be translated differently in French depending on its direct object.
- The noisy channel model is useful for speech recognition but not for machine translation.
- The sentence "A leopard is a cat" entails the sentence "A leopard is a large cat".

no idea pick the 3rd one. The french word joule…

1 point

3. The cosine similarity between the vectors (1,2,0) and (1,1,2) is:

- 0.32
- 1
- 0.55
- 0.80
- 0.5

$$\cos(D,Q) = \frac{|D \cap Q|}{\sqrt{|D||Q|}} = \frac{\sum_{i=1}^{N} d_i q_i}{\sqrt{\sum_{i=1}^{N} d_i^2}\sqrt{\sum_{i=1}^{N} q_i^2}}$$

$$\cos((1,2,0),(1,1,2)) = \frac{1+2+0}{\sqrt{1+4+0}\sqrt{1+1+4}} = \frac{3}{\sqrt{5}\sqrt{6}} = .55$$

---

1 point

4. Which of the following is a subsequence but not a substring of "triangle"?

- ian
- tale
- triple
- tri
- angle

Tricky: read this wrong at first
tale is a subsequence cause t/a/le is there but not a substring of triangle

---

1 point

5. Which of the following statements about the sentences "dogs chase cats" and "cats chase dogs" is true?

- a trigram model will consider the two sentences identical
- a bigram model will consider the two sentences identical
- a bag of words approach will consider the two sentences identical
- a dependency kernel will consider the two sentences identical

BOW words is identical

trigram not same; order matters {dogs chase cats} != { cats chase dogs}
bigram not same; order matters in bigrams {dogs chase, chase cats} != {cats chase, chase dogs}
dependency kernel .. have no idea what this is .

1. The dependency representation of the sentence "Jane has a cat" is the following (with the arrows pointing from parent to child node):

1 / 1 points

2. Which of the following statements is true:

1 / 1 points

3. The cosine similarity between the vectors (1,2,0) and (1,1,2) is:

1 / 1 points

4. Which of the following is a subsequence but not a substring of "triangle"?

1 / 1 points

5. Which of the following statements about the sentences "dogs chase cats" and "cats chase dogs" is true?

1 / 1 points

Week 3 Quiz:

1 point

1. According to the lectures, when using the NLTK implementation of the Lin word similarity method, how similar, on a scale to 1, are the words in each of these pairs? The scores are based on the Brown corpus.

- dog-cat 0.531;

  dog-elephant 0.351
- dog-cat 0.879; dog-elephant 0.531
- dog-cat 0.351;

  dog-elephant 0.531
- dog-cat 0.531;

  dog-elephant 0.209
- dog-cat 0.049;

  dog-elephant 0.048

```
>>> from nltk.corpus import wordnet as wn;
>>> wn.synsets('dog')
[Synset('dog.n.01'), Synset('frump.n.01'), Synset('dog.n.03'), Synset('cad.n.01'), Synset('frank.n.02'), Synset('pawl.n.01'), Synset('andiron.n.01'), Synset('chase.v.01')]
>>> dog=wn.synset('dog.n.01')
>>> cat=wn.synset('cat.n.01')

>>> from nltk.corpus import wordnet_ic
>>> brown_ic=wordnet_ic.ic('ic-brown.dat')
>>> dog.lin_similarity(cat,brown_ic)
0.8768009843733973
>>>

>>> wn.synsets('elephant')
[Synset('elephant.n.01'), Synset('elephant.n.02')]
>>> elephant=wn.synset('elephant.n.01')
>>> dog.lin_similarity(elephant,brown_ic)
0.5312808366793076
```

1 point

2. What is the Levenshtein edit distance between "apples" and "pears"?

Assume the following costs: insertion=1, deletion=1, substitution=1.

- 2
- 5
- 3
- 4
- 1

```
>>> from nltk.metrics.distance import *
>>> edit_distance("car","cars",False)
1
>>> edit_distance("apples","pears",False)
4
>>>
```

1 point

3. When using the Porter stemmer described in class, how does the word "computational" gets stemmed?

- computation
- compute
- computer
- computat
- comput

```
>>> from nltk.stem import PorterStemmer
>>> stemmer=PorterStemmer()
>>> stemmer.stem('computational')
u'comput'
```

1 point

4. Consider the following output of a SRL (Semantic Role Labeling) system:

[X He] [AM-MOD would] [AM-NEG n't] [V accept] [Y anything of value] from [Z those he was writing about] .

The labels X, Y, and Z correspond to which semantic roles?

- X = A0 (acceptor), Y = A1 (thing accepted), Z = A3 (attribute)
- X = A0 (acceptor), Y = A1 (attribute), Z = A2 (accepted-from)
- X = A0 (attribute), Y = A1 (acceptor), Z = A2 (accepted-from)
- X = A0 (attribute), Y = A1 (thing accepted), Z = A2 (accepted-from)
- X = A0 (acceptor), Y = A1 (thing accepted), Z = A2 (accepted-from)

X = A0 Acceptor, Y=thing accepted, Z=A2 accepted-from
Last one

Can we build this? Does this have usefulness?
http://cogcomp.cs.illinois.edu/page/demo_view/SRL
Semantic Role Lableing

1 point

5. When can a matrix S be decomposed as S = U x L x Uinv, where

U is the matrix of eigenvectors of S,

Uinv is the inverse matrix of U, and

L is the diagonal matrix of eigenvalues of S?

Pick the *best* answer below.

- When S is a term x document matrix.
- When U is a diagonal matrix.
- When S maps words from one language to words from another language.
- When U x Uinv = L.
- When S is a square matrix.

when S is square.

Quiz 4:

1 point

1. Which of the following productions is *not* allowed in a Context Free Grammar?

Assume that non-terminals are written in all caps and terminals are in lowercase

- S -> A B c d
- S -> S S
- S -> a b
- A B -> S
- S -> A B

S->S S; the recursion is not allowed

1 point

2. When parsing the sentence "The child saw the cake" using a shift-reduce parser, what are the first five operations that will be performed?

Assume that the parser will use the grammar from the first few slides of Unit 04.03.

- reduce shift reduce shift reduce
- shift reduce shift reduce reduce
- shift shift reduce shift reduce
- shift shift shift shift shift
- shift reduce reduce shift reduce

1
point

3. Consider the CKY table from Unit 04.03. According to the contents of the final version of the table from slides 28-47, how many parses does the sentence "the child ate the cake with the fork" have?

- 2
- 4
- 0
- 1
- 3

1
point

4. Which of the following statements about syntactic constituents is false?

- Constituents are continuous.
- Each word is a constituent.
- Constituents are non-crossing.
- Constituents cannot be nested.
- If two constituents share one word, then one of them must completely contain the other.

1
point

5. In the Earley parser example (Unit 04.04), which of the following chart entries is *not* part of the final parse for the sentence "Take this book".

- [0:3] VP -> VP * PP
- [0 1] VP -> V *
- [0:3] S -> VP *
- [0:1] 'take'
- [0:3] VP -> V NP *

\

Week 2/3 Materials:

Definition: Morphology, where words are composed of smaller morphemes. A morpheme for dogs is dog and s. Or drinkable = drink + able. Ness, able, re, un, etc…

Derivational Morpohology: compose words from other words. Morphemic rules can be used recursively. People understand what these mean. undoable has 2 meanings but unbelievable has 1 meaning. Undoable can be unable to be done or able to be undone.

Inflectional Morpoholgy: in different languages.

Morphological Examples: reduplication;

Morphological Similarity: use stemming to find a base word and use base word to compute similarity. Some words like computation, computable become comput which are the same base word.

http://text-processing.com/demo/stem
http://tartarus.org/martin/PorterStemmer

example of porter stemmed words:

```
>>> from nltk.stem.porter import *
>>> stemmer=PorterStemmer()
>>> stemmer.stem("construction")
u'construct'
>>> stemmer.stem("increasing")
u'increas'
>>> stemmer.stem("unexplained")
u'unexplain'
>>> stemmer.stem("differentiable")
u'differenti'
```

2008 there was a problem with NACLO in 2008 with stemming.

Semantics: meaning of word. 2 types of semantics:

1) Lexical Semantics:
    - hypernym; car is hypernym of prius.. more general word
    - antonym
    - synonyms
    - meronym: tire is meronym of car…
    - also words which have multiple meanings
    - morphological simarity
    - spelling similarity (edit distance)

2) Compositional Semantics: how to understand meaning of sentence based on words as components

New Topic Text Similarity:
There are different types of text similarity. Express related concepts in different ways. A similarity system should detect these occurrences.

How to model text similarity computationally? Define text similarity using humans as ground truth first.

Human judgements of similarity ie. tiger/cat = 7.35, etc.. there are tables for these. Source is 2014 paper from Felix Hill, 999 words, e.g. horse/mare = 8.33. These kind of datasets can be used to train. Mikolov uses word2vec; can compute words most similar to france.. with number.

let's look at definitions for different types of similarity:
1) morphological similarity: respect/respectful; use stemmer to figure out if same, we want to be able to identify using stemming (are abbreviations here?), base form, inflected forms. Stemming a word converts it to a base form. The measure os a word is # of syllables in word. If the measure > 0; stemmer stems the syllable.

2) Spelling Similarity: Edit distance. Delete/Insert/Substitution example.Uses Dynamic programming

http://www.nltk.org/_modules/nltk/metrics/distance.html

3) synonomy: synonyms similar meaning talkative/chatty
3a) hypernymy: deer is hypernym for elk
3b) hyponymy: inverse of hypernymy:
3c) membership meronymy: flock includes sheep or birds or implies things which flock
4) homophony: raise/rays/raze (sounds similar but meaning different), deer is hy
5) semantic similariity: cat/tabby
Differenet words can have same meaning e.g tepid/lukewarm
Polysemy: words have multiple senses; book can refer to literary work/stack of pages, record of bets like in bookmaker.etc… The same word can have multiple parts of speech
a SYNSET (used in nltk) is used to group together group of words. If a word is polysemy then it belongs in multiple synsets
6) sentence similarity: paraphrases of sentences; they may be similar
7) document similarity: different words but documents can be similar
8) cross lingual similarity

Wordnet/handtuned database of words. Tree like. Provides hypernymy relationships.

The word bar has 11 senses as a noun adn 4 senses as a verb.
noun: bar(drink), bar to hold back, etc…

MESH for medical subject headings:
http://www.nlm.nih.gov/mesh/MBrowser.html

Wordnet distance based on links between words. sim(w1,w2)=pathlength(w1,w2) better to take log, sim(w1,w2)= log(pathlength(w1,w2))

Version 3 from Resnik means based on corpus; looks at:
sim(w1,w2)=-logP(LCS(w1,w2))

LCS=lowest common subsumer (from wordnet, common parrent for w1,w2)

Version 4 from Dekang Lin augments wordnet w probabilities
IC(c) = -logP(c)

Sim(w1,w2) = 2*logP(LCS(w1,w2)) / (logP(w1) + logP(w2))

sim('Hill','Coast') = 2*logP('Geological-Formation')/(logP('Hill') + logP('Coast'))=0.59

Dimension Reduction:

SVD/LSI:

if S is square; can decompose to UAU^-1 where U is matrix of Eigenvectors, A is diagonal of eigenvalues

if not square use SVD; A=USV(T) where U is matrix of orthogonal eigenvectors AA(T),
V is matrix of orthogonal eigenvectors of A(T)A, S is the eigenvalues of A(T)A.

Matlab use [U,S,V] = svd(A)

For documents: model as bipartite graph

Add example for document similarity/terms

Approximage SVD:
A* = US*V(T) where S* is the largest eigenvalues

rank-2 approximation; keeps 2 largest eigenvalues; (example)
show graph

if A is document to Term matrix. what is:
AAT is term/term similarity; not clear he uses A*A'
ATA is documet/document (week 3 end)