

Language Models

Introduction to Natural Language Processing
Computer Science 585—Fall 2009
University of Massachusetts Amherst

David Smith

Language Models

- Qualitative observations of words
- N-gram generation
- N-gram language models
- Naive Bayes classifiers
- Applications of language models

Defining Words

How to Write a Dictionary

† **dictio'narian.** *Obs. rare.* [f. as prec. + -AN.]
The maker of a dictionary; a lexicographer.
1846 WORCESTER cites DR. DAWSON.

† **dictionary.** *Obs. rare.* [f. next + -IST.] The
maker of a dictionary.
1617 COLLINS *Def. Bp. Ely* II. vi. 238 One of the
Dictionaryists aforementioned [viz. Budæus, Crispinus] quotes
the place.

dictionary ('dɪksənəri). [ad. med.L.
dictiōnārium or *dictiōnārius* (sc. *liber*) lit. 'a
repertory of *dictiōnēs*, phrases or words' (see
DICTION) in F. *dictionnaire* (R. Estienne 1539),
It. *dizionario*, Sp. *diccionario*.]

1. a. A book dealing with the individual words
of a language (or certain specified classes of
them), so as to set forth their orthography,
pronunciation, signification, and use, their
synonyms, derivation, and history, or at least
some of these facts: for convenience of
reference, the words are arranged in some stated
order, now, in most languages, alphabetical; and
in larger dictionaries the information given is
illustrated by quotations from literature; a
word-book, vocabulary, or lexicon.

Dictionaries proper are of two kinds: those in which the
meanings of the words of one language or dialect are given
in another (or, in a polyglot dictionary, in two or more
languages), and those in which the words of a language are
treated and illustrated in this language itself. The former
were the earlier.

Dictionary was used c 1225 by Joannes de Garlandia, a
native of England, as the title of a collection of Latin
vocabulary, arranged according to their subjects, in sentences,
for the use of learners; e.g.

In herbo magistri Iohannis sunt herbe scilicet iste: salvia,

How to Write a Dictionary

† **dictio'narian.** *Obs. rare.* [f. as prec. + -AN.]
The maker of a dictionary; a lexicographer.
1846 WORCESTER cites DR. DAWSON.

† **dictionary**. *Obs. rare.* [f. next + -IST.] The
maker of a dictionary.
1617 COLLINS *Def. Bp. Ely* II. vi. 238 One of the
Dictionaryists aforementioned [viz. Budæus, Crispinus] quotes
the place.

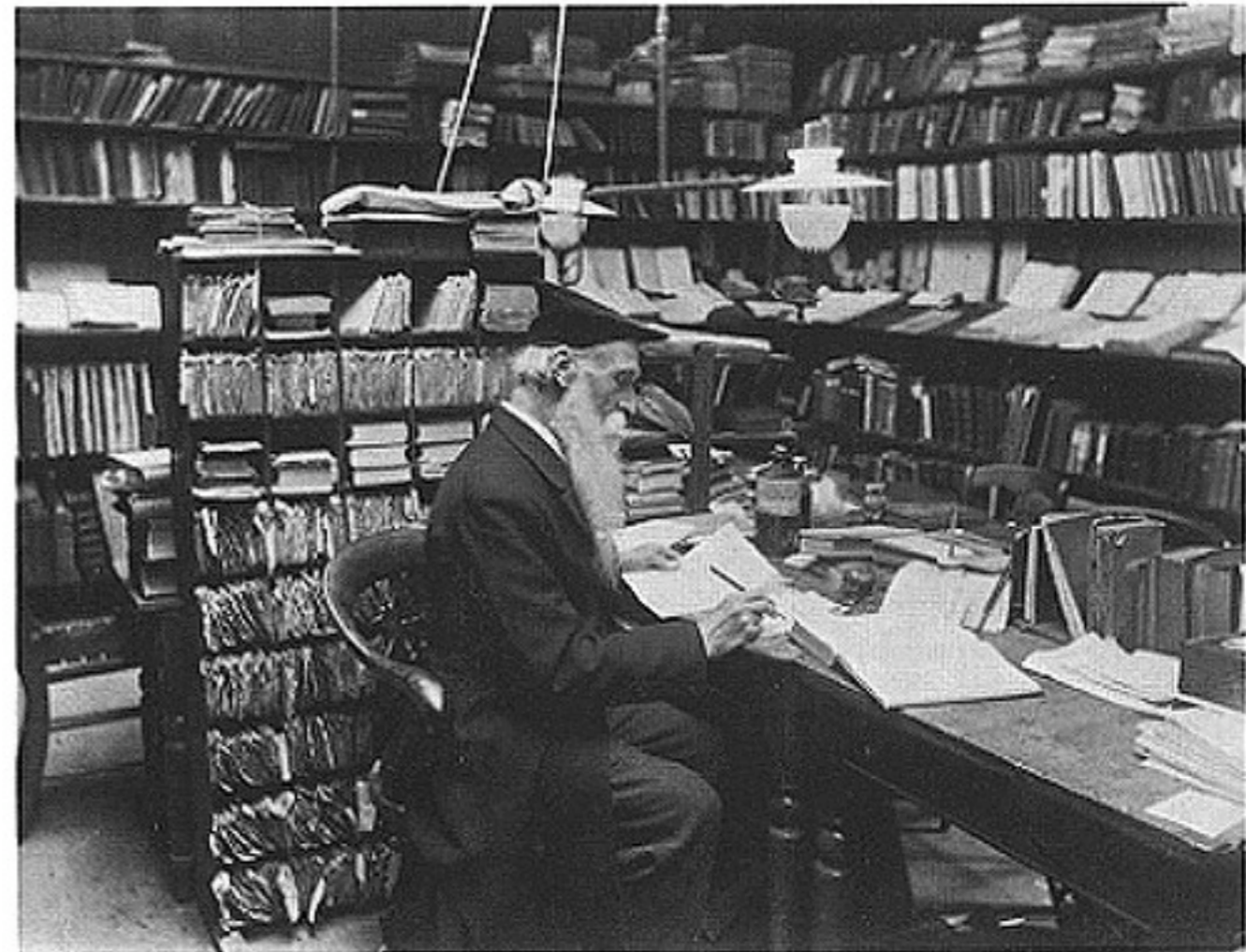
dictionary ('dɪksənəri). [ad. med.L.
dictiōnārium or *dictiōnārius* (sc. *liber*) lit. 'a
repertory of *dictiōnēs*, phrases or words' (see
DICTION) in F. *dictionnaire* (R. Estienne 1539),
It. *dizionario*, Sp. *diccionario*.]

1. a. A book dealing with the individual words
of a language (or certain specified classes of
them), so as to set forth their orthography,
pronunciation, signification, and use, their
synonyms, derivation, and history, or at least
some of these facts: for convenience of
reference, the words are arranged in some stated
order, now, in most languages, alphabetical; and
in larger dictionaries the information given is
illustrated by quotations from literature; a
word-book, vocabulary, or lexicon.

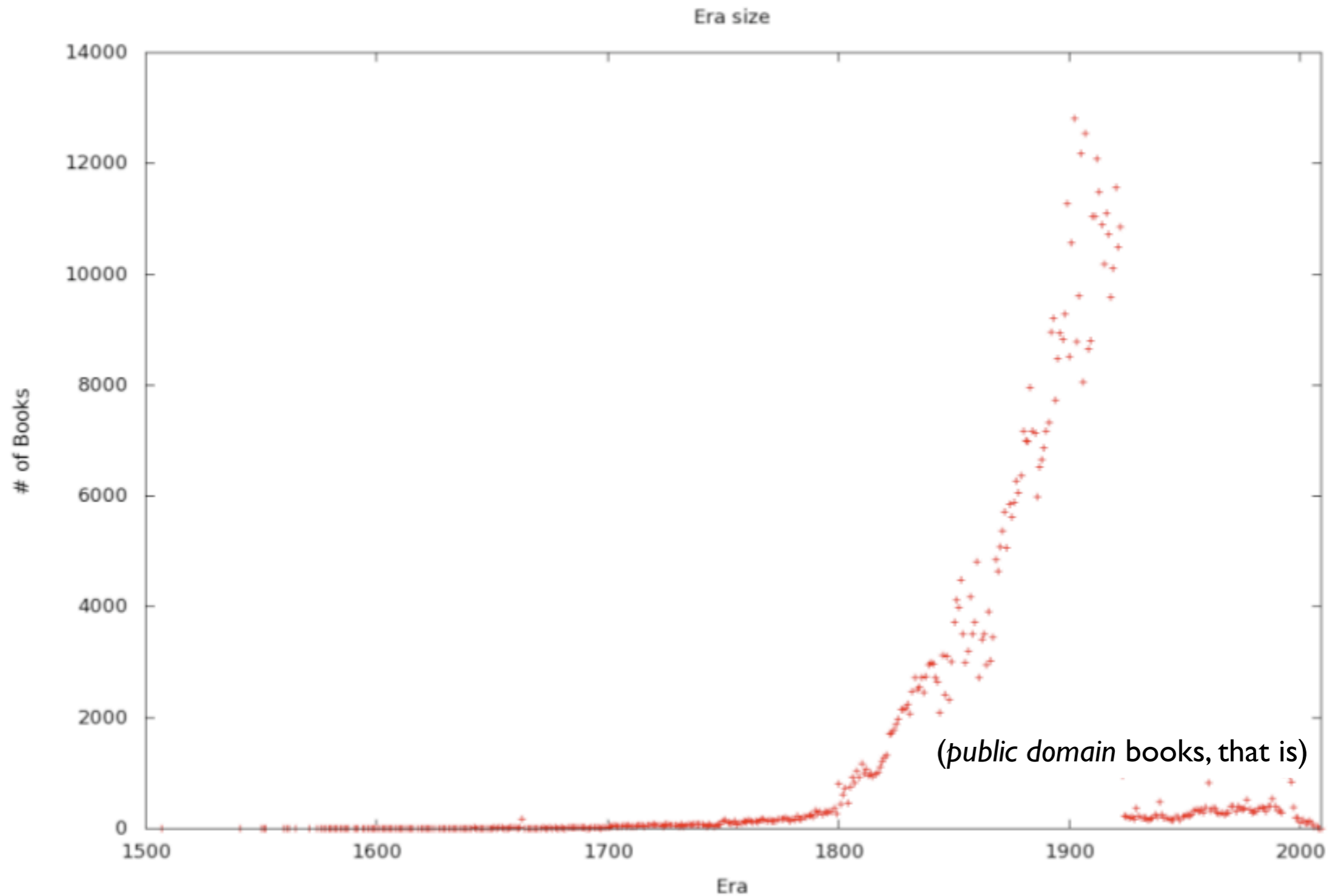
Dictionaries proper are of two kinds: those in which the
meanings of the words of one language or dialect are given
in another (or, in a polyglot dictionary, in two or more
languages), and those in which the words of a language are
treated and illustrated in this language itself. The former
were the earlier.

Dictionary was used c 1225 by Joannes de Garlandia, a
native of England, as the title of a collection of Latin
vocables, arranged according to their subjects, in sentences,
for the use of learners; e.g.

In herbo magistri Iohannis sunt herbe scilicet iste: salvia,



Book Publication Dates



Concordances

```
text1.concordance("match")
```

```
Displaying 9 of 9 matches:
```

```
t in the seventh heavens . Elsewhere match that bloom of theirs , ye cannot , s  
ey all stand before me ; and I their match . Oh , hard ! that to fire others ,  
h , hard ! that to fire others , the match itself must needs be wasting ! What  
so sweet on earth -- heaven may not match it !-- as those swift glances of war  
end ; but hardly had he ignited his match across the rough sandpaper of his ha  
utting the lashing of the waterproof match keg , after many failures Starbuck c  
asks heaped up in him and the slow - match silently burning along towards them  
followed by Stubb ' s producing his match and igniting his pipe , for now a re  
aspect , Pip and Dough - Boy made a match , like a black pony and a white one
```

```
text2.concordance("match")
```

```
Displaying 15 of 15 matches:
```

```
isregarded her disapprobation of the match . Mr . John Dashwood told his mother  
ced of it . It would be an excellent match , for HE was rich , and SHE was hand  
you have any reason to expect such a match ." " Don ' t pretend to deny it , be  
ry much . But mama did not think the match good enough for me , otherwise Sir J  
on ' t we all know that it must be a match , that they were over head and ears  
ght . It will be all to one a better match for your sister . Two thousand a yea  
the other an account of the intended match , in a voice so little attempting co  
end of a week that it would not be a match at all . The good understanding betw  
d with you and your family . It is a match that must give universal satisfactio  
le on him a thousand a year , if the match takes place . The lady is the Hon .  
before , that she thought to make a match between Edward and some Lord ' s dau  
e , with all my heart , it will be a match in spite of her . Lord ! what a taki  
certain penury that must attend the match . His own two thousand pounds she pr  
man nature . When Edward ' s unhappy match takes place , depend upon it his mot  
m myself , and dissuade him from the match ; but it was too late THEN , I found
```

Generating Text

What Doesn't Belong?

The dignity of her own size , or elbowed in old Gomorrah ? No , and all things called the world ' s talking about carrying a sudden , the skies , in bright Greece , ere long time . Didn ' s the Nantuckois , whose skeleton at street being the first faintest glimmering of the Indian and exasperating stories ? The ship , that some reason of fine spirits were pulling towards the smoke in Persia , among the high hump was not a second iron bolt in genteel comedies , so that their tales , because the bottom

The only upon a great , the reasons I believed possible and wishes of tomorrow . No altered by an altogether ." But it into effect . Just as important consideration and ideas of the year after the powers , and are exposed to it is the future . It ought always changing times was east , stronger than any armed balance placed it will also imposes . There was effected the oaths which government and the midst of our fellow - nourished . Utterly , and they had a precious gift of law . I shall be assured confidence only

The animosities of universal forces of six months of the use every other generation of that the 14th day their chosen with perils , is still it was too clear warrant the sound morals and other States have consumed by fearful degree of our republican government by experience will on the Constitution itself . We find impunity in the world . It will be preserved this occasion , and their hearts and settled disposition to produce it . Those which the conflicts and now universal forces and stern performance of the States in the South America . I bring to promote

N-Gram Models

```
print tokenwrap(NgramModel(2,text4).generate(100, ["The"]))
```

The results have secured and reason , and establish free nations should be necessary implements of negotiation , because we have this question of all the best way . Standing here to count on conditions under the case of essential democracy they can truly grateful and noblest structure of the 30th day of silent prayer . They wasted no place called on that a spiral that we will be the American Dream alive . But there , indeed , more minute exposition of a grateful country and rapid transit throughout the people are privileged class among them , when , I

```
print tokenwrap(NgramModel(3,text4).generate(100, ["The"]))
```

The circumstances were , perhaps , our best effort , foreign or domestic , as in the executive departments concerned in its objects . Upon one point there is less occasion for an increase of race feeling among our people to suggest regional groupings of free government , and its amendments . All the constitutional expositions of Webster would courageously avow and defend their real character , however , have altered their aspect as we can count on an ancient hope that government is that Congress in 1890 . The striking incidents of your leading hopes is doomed to disappointment ,

```
print tokenwrap(NgramModel(4,text4).generate(100, ["The"]))
```

The experiment has been made brighter with the presence here of one who , for a minority of their own debasement and ruin . Let us , then , is the most humane , the most advanced , and in power as this nation has seen during the century , the number of men to guard this Heaven - favored land against the mischiefs which without His guidance might arise from an unwise public policy . With a good conscience our only sure reward , with history the final judge of our deeds , that we are the heirs of the

Language Model Demo

Applications of LM

- Language identification
- Text categorization
- Spelling correction
- Speech recognition
- Machine translation
- Grading essays (!)