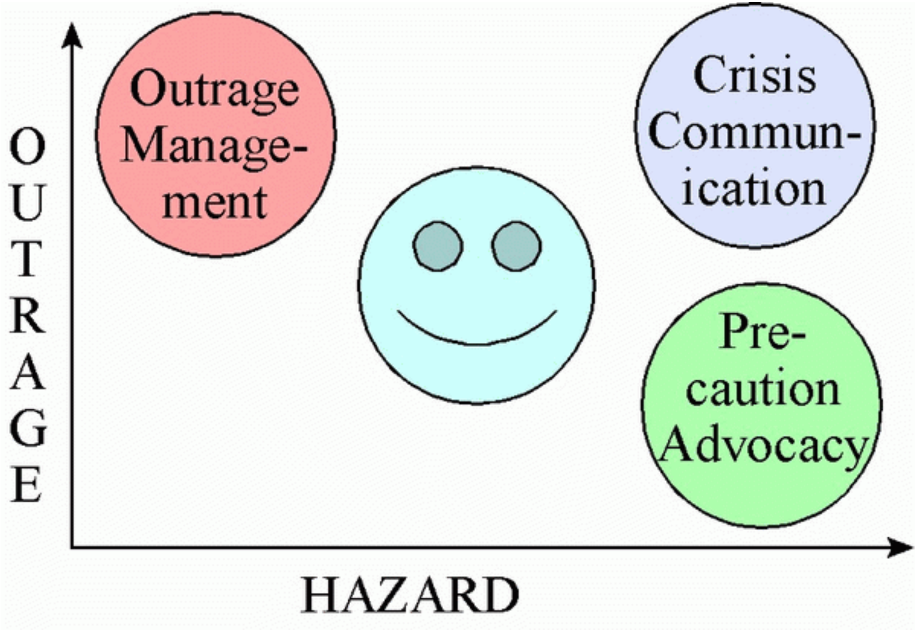
Sackler Forum 2014

Eric Grosse   ehg@google.com

notice and consent

communication of risk

- psandman.com



while:

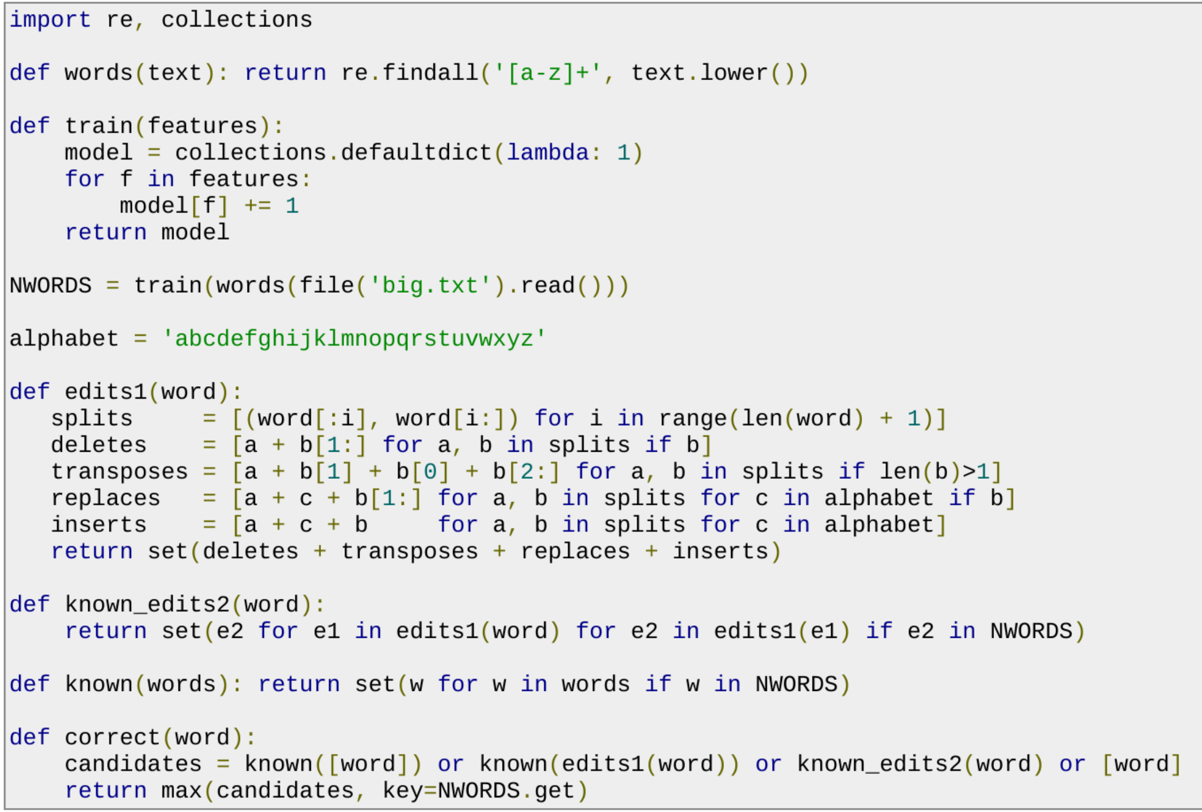
- logged-in
- logged-out
- third party

“pre-specified, limited use”

norvig.com/
spell-correct.html

simple+bigdata >
sophisticated

rare languages



```
import re, collections

def words(text): return re.findall('[a-z]+', text.lower())

def train(features):
    model = collections.defaultdict(lambda: 1)
    for f in features:
        model[f] += 1
    return model

NWORDS = train(words(file('big.txt').read()))

alphabet = 'abcdefghijklmnopqrstuvwxyz'

def edits1(word):
   splits     = [(word[:i], word[i:]) for i in range(len(word) + 1)]
   deletes    = [a + b[1:] for a, b in splits if b]
   transposes = [a + b[1] + b[0] + b[2:] for a, b in splits if len(b)>1]
   replaces   = [a + c + b[1:] for a, b in splits for c in alphabet if b]
   inserts    = [a + c + b     for a, b in splits for c in alphabet]
   return set(deletes + transposes + replaces + inserts)

def known_edits2(word):
    return set(e2 for e1 in edits1(word) for e2 in edits1(e1) if e2 in NWORDS)

def known(words): return set(w for w in words if w in NWORDS)

def correct(word):
    candidates = known([word]) or known(edits1(word)) or known_edits2(word) or [word]
    return max(candidates, key=NWORDS.get)
```

controls on use - guiding principles

1. Benefit user, or defend others.
2. Without security there is no privacy.
3. Let user inspect, export, delete.   Maybe edit.
4. Be *very* skeptical of “anonymization”.

unlinked persona, “hide in plain sight”

controls on collection

Don’t give up too soon!   pervasive crypto, permission pushback

"If you don't want it collected, don't send it."   location

But: car license plate cameras, cell tower triangulation

It’s harder than you may think...

“client ID mechanisms” goo.gl/pyXYv7

explicitly assigned client id

cookies, Flash, Silverlight, HTML5

cache metadata   (Mod-Time, AppCache manifest)

compression dictionary

low-level protocols   (client keys; DNS cache)

incidental fingerprinting   (EFF Panopticlick)

may be only a few bits, but combined with others is enough

(User-Agent, clock skew, DLL, fonts, network ports)

user behavior

language pref, time of data of work

accelerometer

non-default privacy settings

Really hard to hide in plain sight against today’s best (or unscrupulous) actors

ultimate control: trust

Give value for what you take. Be reciprocal.

If you abuse it, you will lose it.

p.s. Let’s maintain respectful dialog.

housespecial.com/momentum