

✓ Nama Kelompok: Finland

5210411257 - Ridwan Halim

5210411277 - Taufiq Dwi Purnomo

5210411285 - Guillianio Rasyid

5210411286 - M. Sidiq Pramono

[link colab](#)

```
1 import nltk
2 from nltk.corpus import brown, reuters, inaugural, gutenberg
3
4 nltk.download('brown')
5 nltk.download('reuters')
6 nltk.download('inaugural')
7 nltk.download('gutenberg')
8 nltk.download('punkt')
```

```
[nltk_data] Downloading package brown to /root/nltk_data...
[nltk_data] Package brown is already up-to-date!
[nltk_data] Downloading package reuters to /root/nltk_data...
[nltk_data] Package reuters is already up-to-date!
[nltk_data] Downloading package inaugural to /root/nltk_data...
[nltk_data] Package inaugural is already up-to-date!
[nltk_data] Downloading package gutenberg to /root/nltk_data...
[nltk_data] Package gutenberg is already up-to-date!
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Unzipping tokenizers/punkt.zip.
True
```

✓ Tugas

✓ 1. Cobalah corpus brown dengan menggunakan brown.words()

a. Categories: fiction, mystery, humor, romance

b. Files: cd12, cb02, cj19, cm01

```
1 for file in brown.fileids(['fiction', 'mystery', 'humor', 'romance']):
2     print(brown.words(file))
```

```
['Thirty-three', 'Scotty', 'did', 'not', 'go', 'back', ...]
['Where', 'their', 'sharp', 'edges', 'seemed', ...]
['Mickie', 'sat', 'over', 'his', 'second', ...]
['The', 'Bishop', 'looked', 'at', 'him', 'coldly', ...]
['Payne', 'dismounted', 'in', 'Madison', 'Place', ...]
['With', 'a', 'sneer', 'the', 'man', 'spread', ...]
['If', 'the', 'crummy', 'bastard', 'could', 'write', ...]
['Rousseau', 'is', 'so', 'persuasive', 'that', ...]
['It', 'was', 'the', 'first', 'time', 'any', 'of', ...]
['That', 'summer', 'the', 'gambling', 'houses', 'were', ...]
['Standing', 'in', 'the', 'shelter', 'of', 'the', ...]
['She', 'was', 'a', 'child', 'too', 'much', 'a', ...]
['In', 'the', 'dim', 'underwater', 'light', 'they', ...]
['He', 'brought', 'with', 'him', 'a', 'mixture', 'of', ...]
['Beth', 'was', 'very', 'still', 'and', 'her', ...]
['The', 'red', 'glow', 'from', 'the', 'cove', 'had', ...]
['Burly', 'leathered', 'men', 'and', 'wrinkled', ...]
['She', 'was', 'getting', 'real', 'dramatic', '.', ...]
['There', 'was', 'one', 'fact', 'which', 'Rector', ...]
['She', 'concluded', 'by', 'asking', 'him', 'to', ...]
['Beckworth', 'handed', 'the', 'pass', 'to', 'the', ...]
['I', 'would', 'not', 'want', 'to', 'be', 'one', 'of', ...]
['It', 'was', 'not', 'as', 'though', 'she', 'noted', ...]
['His', 'eyes', 'were', 'old', 'and', 'they', 'never', ...]
['He', 'was', 'in', 'his', 'mid-fifties', 'at', 'this', ...]
['But', 'they', 'all', 'said', 'No', '...', ...]
['But', 'tell', 'me', 'doctor', '...', ...]
['Going', 'downstairs', 'with', 'the', 'tray', '...', ...]
['Was', 'it', 'love', 'I', 'had', 'no', ...]
['There', 'were', 'thirty-eight', 'patients', 'on', ...]
['I', 'was', 'giving', 'the', 'parked', 'cars', 'the', ...]
['The', 'fat', 'man', 'said', 'All', 'we', ...]
['His', 'jowls', 'were', 'spiked', 'by', 'barbs', 'of', ...]
```

```

['Then', 'he', 'turned', 'the', 'telephone', 'over', ...]
['Eight', ' ', 'nine', 'steps', 'above', 'him', ' ', ...]
['He', 'put', 'in', 'a', 'call', 'to', 'Cunningham', ...]
[''', 'Well', '""', '--', 'said', 'Mr.', 'Skyros', ...]
["Maude's", 'long', 'nose', 'unexpectedly', 'wrinkled', ...]
[''', 'Not', 'since', 'last', 'night', ' ', 'I', ...]
[''', 'Right', '""', ' ', 'said', 'the', ...]
['Andy', 'did', 'not', 'see', 'the', 'newspapers', ...]
['His', 'son', 'watched', 'until', 'he', 'got', 'as', ...]
['A', 'man', 'with', 'a', 'sketch', 'pad', 'in', ...]
[''', 'Through', 'a', 'door', 'conveniently', ...]
[''', 'Dammit', ' ', 'Phil', ' ', 'are', 'you', ...]
['The', 'safe', 'at', 'Ingleside', 'District', ...]
[''', 'She', 'says', 'she', 'has', 'to', 'finish', ...]
['Slowly', 'he', 'pulled', 'out', 'the', 'hand', ...]
['Harbor', 'Point', 'sticks', 'out', 'into', 'the', ...]
['But', 'the', 'police', 'have', 'dropped', 'the', ...]
['In', 'good', 'time', 'I', 'shall', 'get', 'to', ...]
["That's", 'what', 'the', 'man', 'had', 'said', ' ', ...]
['About', 'halfway', 'back', 'Pops', 'groped', ...]
['They', 'neither', 'liked', 'nor', 'disliked', 'the', ...]
['People', 'came', 'in', 'and', 'out', 'all', ...]
[''', 'Thrifty', 'of', 'her', 'to', 'use', 'it', 'up', ...]
[''', 'He', 'must', 'have', 'forgiven', 'me', '""', ...]
["That 'Memoranda' family had come from"]

```

```
1 brown.words(['cd12', 'cb02', 'cj19', 'cm01'])
```

```
['Much', 'more', 'than', 'shelter', ' ', 'housing', ...]
```

✓ 2. Cobalah corpus brown dengan menggunakan dengan isi, kode tahun '1993'

a. `inaugural.raw()`

b. `inaugural.words()`

c. `inaugural.sents()`

d. `inaugural.paras()`

```

1 for i in inaugural.fileids():
2     if '1993' in i:
3         raw = inaugural.raw(i)
4         words = inaugural.words(i)
5         sents = inaugural.sents(i)
6         paras = inaugural.paras(i)

```

```
1 print(raw)
```

My fellow citizens, today we celebrate the mystery of American renewal. This ceremony is held in the depth of winter, but by the work of our hands and the power of our minds, we have begun to see the light of a new day. Though we march to the music of our time, our mission is timeless. Each generation of American's must define what it means to be an American. Today, a generation raised in the shadows of the Cold War assumes new responsibilities in a world warmed by the sunshine of freedom, when George Washington first took the oath I have just sworn to uphold, news travelled slowly across the land by horseback, and across the land by the power of the press. We earn our livelihood in America today in peaceful competition with people all across the Earth. Profound and powerful forces are at work in the world, and we must be ready to meet them. We know we have to face hard truths and take strong steps, but we have not done so. Instead we have drifted, and that drifting has cost us dearly. Our democracy must be not only the envy of the world but the engine of our own renewal. There is nothing wrong with America that cannot be fixed. And so today we pledge an end to the era of deadlock and drift, and a new season of American renewal has begun.

To renew America we must be bold. We must do what no generation has had to do before. We must invest more in our own people, in their education, in their health, in their homes. We must provide for our nation the way a family provides for its children. Our founders saw themselves in the light of posterity. We must do the same. It is time to break the bad habit of expecting something for nothing: from our government, or from each other. Let us all take more responsibility for our country. Americans deserve better, and in this city today there are people who want to do better, and so I say to all of you here, let us resolve to do better. To renew America we must meet challenges abroad, as well as at home. There is no longer a clear division between what is foreign and what is domestic. Together with our friends and allies, we will work together to shape change, lest it engulf us. When our vital interests are challenged, we will stand together. My fellow Americans, you, too, must play your part in our renewal. I challenge a new generation of YOUNG Americans to a season of service. And so, my fellow Americans, as we stand at the edge of the 21st Century, let us begin anew, with energy and hope, with faith and determination. Thank you, and God bless you all.

```

1 print(words)

['My', 'fellow', 'citizens', ',', 'today', 'we', ...]

1 print(sents)

[['My', 'fellow', 'citizens', ',', 'today', 'we', 'celebrate', 'the', 'mystery', 'of', 'American', 'renewal', '.'], ['This', 'cerem
1 print(paras)

[['My', 'fellow', 'citizens', ',', 'today', 'we', 'celebrate', 'the', 'mystery', 'of', 'American', 'renewal', '.'], ['This', 'ceren

```

3. Tampilkan (5) kategori dan (10) fields pada corpus reuters!

```

1 reuters.categories()[5]

['acq', 'alum', 'barley', 'bop', 'carcass']

1 reuters.fileids()[10]

['test/14826',
 'test/14828',
 'test/14829',
 'test/14832',
 'test/14833',
 'test/14839',
 'test/14840',
 'test/14841',
 'test/14842',
 'test/14843']

1 for cat in reuters.categories()[5]:
2     print(reuters.fileids(cat)[10])

['test/14843', 'test/14852', 'test/14865', 'test/14888', 'test/14900', 'test/14907', 'test/14909', 'test/14921', 'test/14932', 'test/
['test/15368', 'test/16179', 'test/16318', 'test/16320', 'test/16750', 'test/16868', 'test/17698', 'test/17702', 'test/17714', 'test/
['test/15618', 'test/15649', 'test/15676', 'test/15728', 'test/15871', 'test/15875', 'test/15952', 'test/17767', 'test/17769', 'test/
['test/14862', 'test/15543', 'test/15573', 'test/15646', 'test/15793', 'test/16108', 'test/16158', 'test/16226', 'test/16759', 'test/
['test/14858', 'test/15254', 'test/15642', 'test/15798', 'test/16379', 'test/16564', 'test/16784', 'test/18034', 'test/18099', 'test/

```