

lab5

April 7, 2023

Week 06

Student name: Mohammed Obidou

Student_id: 119200016

Student email: Mohammed.obidou@bilgiedu.tr

Navigation

- PART 1 | Test Corpus path
- PART 2 | Create world list corpus
- PART 3 | Create a corpus from URL
- PART 4 | Create corpus from text file
- PART 5 | IBM MODEL

PART 1 | Test corpus path

```
[4]: from nltk import *  
import os, os.path
```

```
[5]: path = os.path.expanduser('~\\nltk_data')
```

```
[6]: path
```

```
[6]: 'C:\\Users\\ayman\\nltk_data'
```

```
[7]: import nltk.data
```

```
[8]: for path in nltk.data.path:  
    print(path)
```

```
C:\\Users\\ayman\\nltk_data  
c:\\Users\\ayman\\AppData\\Local\\Programs\\Python\\Python310\\nltk_data  
c:\\Users\\ayman\\AppData\\Local\\Programs\\Python\\Python310\\share\\nltk_data  
c:\\Users\\ayman\\AppData\\Local\\Programs\\Python\\Python310\\lib\\nltk_data  
C:\\Users\\ayman\\AppData\\Roaming\\nltk_data  
C:\\nltk_data  
D:\\nltk_data  
E:\\nltk_data
```

```
[9]: data=nltk.data.load('a.text', format = 'raw')
[10]: data
[10]: b'meow meow meow'
[11]: data2 = nltk.data.load('b.text', format='raw')
[12]: data2
[12]: b'wof wof wof'
[13]: data3 = nltk.data.load('c.text', format='raw')
data3
[13]: b'A prejudice is an addiction, and
it\xe2\x80\x99s\r\ncontagious\xe2\x80\x94parents infect their children.\r\nAnd
addiction\xe2\x80\x99s obsessive, if a man finds it\r\nndifficult to show his
love to his\r\nson, it may be because his father\r\nescaped with his life from
the village in which\r\nhis own father had just been murdered\r\nnin a pogrom,
his model as a father\r\na man in terror.\r\nBut addiction to such a silence can
be\r\nhealed, as Carl and his son tried to do,\r\nthrough hard work. Workers of
the world,\r\nn unite, we have nothing to lose\r\nbut the death of the earth.\r\n'
```

PART 2 | Create a WordList corpus

```
[14]: from nltk.corpus.reader import *
[15]: readerA = WordListCorpusReader('C:\\Users\\ayman\\OneDrive - Istanbul Bilgi_
↳Universitesi\\Coding-New\\Coding\\WORKSPACE\\PYTHON\\machine_
↳translation\\lab5', 'a.text')
[16]: readerA.words()
[16]: ['meow meow meow']
[17]: readerB = WordListCorpusReader('C:\\Users\\ayman\\OneDrive - Istanbul Bilgi_
↳Universitesi\\Coding-New\\Coding\\WORKSPACE\\PYTHON\\machine_
↳translation\\lab5', 'b.text')
readerB.words()
[17]: ['wof wof wof']
[61]: reader_All = WordListCorpusReader('C:\\Users\\ayman\\OneDrive - Istanbul Bilgi_
↳Universitesi\\Coding-New\\Coding\\WORKSPACE\\PYTHON\\machine_
↳translation\\lab5', ['b.text', 'a.text', 'c.text'])
reader_All.words()
```

```
[61]: ['wof wof wofmeow meow meowA prejudice is an addiction, and it's',
      'contagious-parents infect their children.',
      'And addiction's obsessive, if a man finds it',
      'difficult to show his love to his',
      'son, it may be because his father',
      'escaped with his life from the village in which',
      'his own father had just been murdered',
      'in a pogrom, his model as a father',
      'a man in terror.',
      'But addiction to such a silence can be',
      'healed, as Carl and his son tried to do,',
      'through hard work. Workers of the world,',
      'unite, we have nothing to lose',
      'but the death of the earth.']
```

```
[19]: print("file names are: ",reader_All.fileids())
```

```
file names are: ['b.text', 'a.text', 'c.text']
```

```
[20]: print("output in raw format...")
      print(reader_All.raw())
```

```
output in raw format...
```

```
wof wof wofmeow meow meowA prejudice is an addiction, and it's
contagious-parents infect their children.
And addiction's obsessive, if a man finds it
difficult to show his love to his
son, it may be because his father
escaped with his life from the village in which
his own father had just been murdered
in a pogrom, his model as a father
a man in terror.
But addiction to such a silence can be
healed, as Carl and his son tried to do,
through hard work. Workers of the world,
unite, we have nothing to lose
but the death of the earth.
```

Line Tokenize

```
[21]: from nltk.tokenize import line_tokenize
```

```
[22]: line_tokenize(readerA.raw())
```

```
[22]: ['meow meow meow']
```

```
[23]: line_tokenize(reader_All.raw())
```

```
[23]: ['wof wof wofmeow meow meowA prejudice is an addiction, and it's',  
      'contagious-parents infect their children.',  
      'And addiction's obsessive, if a man finds it',  
      'difficult to show his love to his',  
      'son, it may be because his father',  
      'escaped with his life from the village in which',  
      'his own father had just been murdered',  
      'in a pogrom, his model as a father',  
      'a man in terror.',  
      'But addiction to such a silence can be',  
      'healed, as Carl and his son tried to do,',  
      'through hard work. Workers of the world,',  
      'unite, we have nothing to lose',  
      'but the death of the earth.']
```

```
[24]: # test english corpus
```

```
[25]: from nltk.corpus import *
```

```
[26]: words.fileids()
```

```
[26]: ['en', 'en-basic']
```

```
[27]: len(words.words())
```

```
[27]: 236736
```

```
[28]: len(words.words('en-basic'))
```

```
[28]: 850
```

```
[29]: words.words('en-basic')
```

```
[29]: ['I',  
      'a',  
      'able',  
      'about',  
      'account',  
      'acid',  
      'across',  
      'act',  
      'addition',  
      'adjustment',  
      'advertisement',  
      'after',  
      'again',  
      'against',  
      'agreement',
```

'toe',
'together',
'tomorrow',
'tongue',
'tooth',
'top',
'touch',
'town',
'trade',
'train',
'transport',
'tray',
'tree',
'trick',
'trouble',
'trousers',
'true',
'turn',
'twist',
'umbrella',
'under',
'unit',
'up',
'use',
'value',
'verse',
'very',
'vessel',
'view',
'violent',
'voice',
'waiting',
'walk',
'wall',
'war',
'warm',
'wash',
'waste',
'watch',
'water',
'wave',
'wax',
'way',
'weather',
'week',
'weight',
'well',

```
'west',  
'wet',  
'wheel',  
'when',  
'where',  
'while',  
'whip',  
'whistle',  
'white',  
'who',  
'why',  
'wide',  
'will',  
'wind',  
'window',  
'wine',  
'wing',  
'winter',  
'wire',  
'wise',  
'with',  
'woman',  
'wood',  
'wool',  
'word',  
'work',  
'worm',  
'wound',  
'writing',  
'wrong',  
'year',  
'yellow',  
'yes',  
'yesterday',  
'you',  
'young']
```

-
- remove stop words
 - remove repeated words
 - Create a wordlist corpus from 5 different files such as: books, cars, animals,countries, sports etc.

```
[43]: import contractions  
from tokenizers import *  
import string, re
```

```

[44]: class Preprocessing_token():
    def __init__(self, token):
        self.token = token

    # returns the expanded version of contractions
    def remove_contractions(self, token):
        token = contractions.fix(token.lower())
        return token

    #convert all words to lower case
    def remove_uppercase(self, token):
        token = token.lower()
        return token

    #Remove Punctuation
    def remove_punctuation(self, token):
        token = re.sub('[%s]' % re.escape(string.punctuation), '', token)
        return token

    #Remove Numbers
    def remove_numbers(self, token):
        token = re.sub(r'\d+', '', token)
        return token

    #Remove whitespace
    def remove_whitespace(self, token):
        token = " ".join(token.split()) #split text then join with space between
        ↪ words
        return token

    #remove Emojis
    def remove_emojis(self, token):
        regex_pattern = re.compile(pattern = "["
            u"\U0001F600-\U0001F64F" # emoticons
            u"\U0001F300-\U0001F5FF" # symbols & pictographs
            u"\U0001F680-\U0001F6FF" # transport & map symbols
            u"\U0001F1E0-\U0001F1FF" # flags (iOS)
            "]" + "", flags = re.UNICODE)
        return regex_pattern.sub(r'', token)

    def remove_repeated(self, token):
        ulist = []
        [ulist.append(x) for x in token if x not in ulist]

        token = " ".join(unique_list(token.split()))
        return token

```

```
[45]: text = "My My parents parents parents aren't coming this this weekend. This_
↳isn't a problem for us. We aren't planning a trip now. I ain't gonna attend._
↳MY plans ARE changed, I will rest 10 hours "
```

```
preproc = Preprocessing_token(text)
text_fix = preproc.remove_contractions(text)
test_upper = preproc.remove_uppercase(text)
test_punctuation = preproc.remove_punctuation(text)
test_numbers = preproc.remove_numbers(text)
test_whitespace = preproc.remove_whitespace(text)
test_emojis = preproc.remove_emojis(text)
test_repeated = preproc.remove_repeated(text)
print(f'Original text is: {text}\n\nWithout Contractions: {text_fix}\n\nWithot_
↳upper: {test_upper}\n\nWithout Punctuation: {test_punctuation}\n\nWithout_
↳Numbers: {test_numbers}\n\nWithout Whitespaces: {test_whitespace}\n\nWithout_
↳Emojis: {test_emojis}\n\nWithout Repeated words: {test_repeated}')
```

Original text is: My My parents parents parents aren't coming this this weekend. This isn't a problem for us. We aren't planning a trip now. I ain't gonna attend. MY plans ARE changed, I will rest 10 hours

Without Contractions: my my parents parents parents are not coming this this weekend. this is not a problem for us. we are not planning a trip now. i are not going to attend. my plans are changed, i will rest 10 hours

Withot upper: my my parents parents parents aren't coming this this weekend. this isn't a problem for us. we aren't planning a trip now. i ain't gonna attend. my plans are changed, i will rest 10 hours

Without Punctuation: My My parents parents parents arent coming this this weekend This isnt a problem for us We arent planning a trip now I aint gonna attend MY plans ARE changed I will rest 10 hours

Without Numbers: My My parents parents parents aren't coming this this weekend. This isn't a problem for us. We aren't planning a trip now. I ain't gonna attend. MY plans ARE changed, I will rest hours

Without Whitespaces: My My parents parents parents aren't coming this this weekend. This isn't a problem for us. We aren't planning a trip now. I ain't gonna attend. MY plans ARE changed, I will rest 10 hours

Without Emojis: My My parents parents parents aren't coming this this weekend. This isn't a problem for us. We aren't planning a trip now. I ain't gonna attend. MY plans ARE changed, I will rest 10 hours

Without Repeated words: My parents aren't coming this weekend. This isn't a problem for us. We planning trip now. I ain't gonna attend. MY plans ARE changed, will rest 10 hours

create a wordlist corpus from 5 files: : books, cars, animals,countries, sports

```
[62]: reader = WordListCorpusReader('C:\\Users\\ayman\\OneDrive - Istanbul Bilgi_
↳Universitesi\\Coding-New\\Coding\\WORKSPACE\\PYTHON\\machine_
↳translation\\lab5',['animals.text', 'sports.text', 'cars.text', 'books.text',_
↳'countries.text'])
reader.words()
```

```
[62]: ['Aardvark',
'Aardwolf',
'Abyssinian',
'Abyssinian Guinea Pig',
'Acadian Flycatcher',
'Achrioptera Manga',
'Ackie Monitor',
'Addax',
'Adelie Penguin',
'Admiral Butterfly',
'Aesculapian Snake',
'Affenpinscher',
'Afghan Hound',
'African Bullfrog',
'African Bush Elephant',
'African Civet',
'African Clawed Frog',
'African Fish Eagle',
'African Forest Elephant',
'African Golden Cat',
'African Grey Parrot',
'African Jacana',
'African Palm Civet',
'African Penguin',
'African Sugarcane Borer',
'African Tree Toad',
'African Wild Dog',
'Africanized bee (killer bee)',
'Agama Lizard',
'Agkistrodon Contortrix',
'Agouti',
'Aidi',
'Ainu',
'Airedale Terrier',
'Airedoodle',
'Akbash',
'Akita',
'Akita Shepherd',
'Alabai (Central Asian Shepherd)',
```

'Honda Civic Type R',
'Subaru Impreza WRX STI',
'Mitsubishi Lancer Evolution',
'Ford RS200',
'Audi RS2 Avant',
'Toyota Supra',
'Lexus LFA',
'Porsche 918 Spyder',
'McLaren P1',
'Lamborghini Aventador',
'Bugatti Veyron',
'Ferrari LaFerrari',
'Koenigsegg Agera',
'Pagani Huayra',
'Tesla Model S',
'Quotations from Chairman Mao Tse-tung',
'Don Quixote by Miguel de Cervantes',
'A Tale of Two Cities by Charles Dickens',
'The Lord of the Rings by J.R.R. Tolkien',
'The Alchemist by Paulo Coelho',
'The Da Vinci Code by Dan Brown',
'The Catcher in the Rye by J.D. Salinger',
'Harry Potter and the Philosopher's Stone by J.K. Rowling',
'The Hobbit by J.R.R. Tolkien',
'The Little Prince by Antoine de Saint-Exupéry',
'Alice's Adventures in Wonderland by Lewis Carroll',
'Dream of the Red Chamber by Cao Xueqin',
'And Then There Were None by Agatha Christie',
'She: A History of Adventure by H. Rider Haggard',
'The Lion, the Witch and the Wardrobe by C.S. Lewis',
'The Adventures of Sherlock Holmes by Arthur Conan Doyle',
'The Diary of a Young Girl by Anne Frank',
'The Adventures of Pinocchio by Carlo Collodi',
'The Name of the Rose by Umberto Eco',
'The Three Musketeers by Alexandre Dumas',
'The Tale of Peter Rabbit by Beatrix Potter',
'The Bridges of Madison County by Robert James Waller',
'The Girl with the Dragon Tattoo by Stieg Larsson',
'The Great Gatsby by F. Scott Fitzgerald',
'The Chronicles of Narnia by C.S. Lewis',
'The Hobbit and The Lord of the Rings by J.R.R. Tolkien',
'The Adventures of Tom Sawyer by Mark Twain',
'The Picture of Dorian Gray by Oscar Wilde',
'The War of the Worlds by H.G. Wells',
'The Hound of the Baskervilles by Arthur Conan Doyle',
'The Secret Garden by Frances Hodgson Burnett',
'The Very Hungry Caterpillar by Eric Carle',
'The Wind in the Willows by Kenneth Grahame',

'The Call of the Wild by Jack London',
'The Godfather by Mario Puzo',
'The Old Man and the Sea by Ernest Hemingway',
"The Hitchhiker's Guide to the Galaxy by Douglas Adams",
'The Adventures of Huckleberry Finn by Mark TwainAfghanistan',
'Albania',
'Algeria',
'Argentina',
'Australia',
'Austria',
'Bangladesh',
'Belgium',
'Brazil',
'Canada',
'China',
'Colombia',
'Cuba',
'Denmark',
'Egypt',
'Ethiopia',
'France',
'Germany',
'Greece',
'India',
'Indonesia',
'Iran',
'Iraq',
'Israel',
'Italy',
'Japan',
'Kenya',
'Mexico',
'Netherlands',
'Nigeria',
'North Korea',
'Norway',
'Pakistan',
'Peru',
'Philippines',
'Poland',
'Portugal',
'Russia',
'Saudi Arabia',
'South Africa',
'South Korea',
'Spain',
'Sweden',

```
'Switzerland',
'Thailand',
'Turkey',
'Ukraine',
'United Kingdom',
'United States of America',
'Vietnam']
```

PART 3 | Create a corpus from URL

```
[47]: # create corpus from webs
```

```
[48]: from urllib import *
```

```
[49]: url = 'https://en.wikipedia.org/wiki/Atomic_physics'
```

```
[50]: response =request.urlopen(url)
```

```
[51]: raw =response.read()
```

```
[52]: len(raw)
```

```
[52]: 134450
```

```
[53]: print(raw)
```

```
b'<!DOCTYPE html>\n<html class="client-nojs vector-feature-language-in-header-
enabled vector-feature-language-in-main-page-header-disabled vector-feature-
language-alert-in-sidebar-enabled vector-feature-sticky-header-disabled vector-
feature-page-tools-pinned-disabled vector-feature-toc-pinned-enabled vector-
feature-main-menu-pinned-disabled vector-feature-limited-width-enabled vector-
feature-limited-width-content-enabled vector-feature-zebra-design-disabled
vector-feature-page-tools-enabled" lang="en" dir="ltr">\n<head>\n<meta
charset="UTF-8"/>\n<title>Atomic physics -
Wikipedia</title>\n<script>document.documentElement.className="client-js vector-
feature-language-in-header-enabled vector-feature-language-in-main-page-header-
disabled vector-feature-language-alert-in-sidebar-enabled vector-feature-sticky-
header-disabled vector-feature-page-tools-pinned-disabled vector-feature-toc-
pinned-enabled vector-feature-main-menu-pinned-disabled vector-feature-limited-
width-enabled vector-feature-limited-width-content-enabled vector-feature-zebra-
design-disabled vector-feature-page-tools-enabled";(function(){var
cookie=document.cookie.match(/(?:^|;)/);
)enwikimwclientprefs=([^\;]+)/);if(cookie){var featureName=cookie[1];document.doc
umentElement.className=document.documentElement.className.replace(featureName+'
-enabled','featureName+'-disabled\');}}());RLCONF={"wgBreakFrames":false,"wgSep
aratorTransformTable":["",""],"wgDigitTransformTable":["",""],"wgDefaultDateForm
at":"dmy","wgMonthNames":["","January","February","March","April","May","June","
July","August","September","October","November","\nDecember"],"wgRequestId":"b87
21b30-92df-4a42-ab50-6f28749ba5fc","wgCSPNonce":false,"wgCanonicalNamespace":"","
```

```
[64]: from bs4 import BeautifulSoup
```

```
[67]: quantom_physics
```

72

Atomic Spectra. Cambridge University Press. ISBN\xa0978-0-521-09209-8.\nCowan, Robert D. (1981). The Theory of Atomic Structure and Spectra. University of California Press. ISBN\xa0978-0-520-03821-9.\nLindgren, I. & Morrison, J. (1986). Atomic Many-Body Theory (Second\xa0ed.). Springer-Verlag. ISBN\xa0978-0-387-16649-0.\nReferences[edit]\n\n\n^ Demtröder, W. (2006). Atoms, molecules and photons\xa0: an introduction to atomic-, molecular-, and quantum-physics. Berlin: Springer. ISBN\xa0978-3-540-32346-4. OCLC\xa0262692011.\n\n\nExternal links[edit]\n\n\nWikimedia Commons has media related to Atomic physics.\n\nMIT-Harvard Center for Ultracold Atoms\nStanford QFARM Initiative for Quantum Science & Engineering\nJoint Quantum Institute at University of Maryland and NIST\nAtomic Physics on the Internet\nJILA (Atomic Physics)\nORNL Physics Division\nvteBranches of physicsDivisions\nPure\nApplied\nEngineering\nApproaches\nExperimental\nTheoretical\nComputational\nClassical\nClassical mechanics\nNewtonian\nAnalytical\nCelestial\nContinuum\nAcoustics\nClassical electromagnetism\nClassical optics\nRay\nWave\nThermodynamics\nStatistical\nNon-equilibrium\nModern\nRelativistic mechanics\nSpecial\nGeneral\nNuclear physics\nQuantum mechanics\nParticle physics\nAtomic, molecular, and optical physics\nAtomic\nMolecular\nModern optics\nCondensed matter physics\nInterdisciplinary\nAstrophysics\nAtmospheric physics\nBiophysics\nChemical physics\nGeophysics\nMaterials science\nMathematical physics\nMedical physics\nOcean physics\nQuantum information science\nRelated\nHistory of physics\nNobel Prize in Physics\nPhysics education\nTimeline of physics discoveries\n\nAuthority control National\nGermany\nJapan\nCzech Republic\nOther\nEncyclopedia of Modern Ukraine\n\n\n\nRetrieved from "https://en.wikipedia.org/w/index.php?title=Atomic_physics&oldid=1147433735"\nCategories: Atomic physicsAtomic, molecular, and optical physicsHidden categories: Articles with short descriptionShort description is different from WikidataArticles lacking in-text citations from September 2015All articles lacking in-text citationsArticles containing Sanskrit-language textLang and lang-xx code promoted to ISO 639-1Articles containing German-language textArticles containing Italian-language textArticles containing Russian-language textArticles containing Macedonian-language textCommons category link from WikidataArticles with GND identifiersArticles with NDL identifiersArticles with NKC identifiersArticles with EMU identifiers\n\n\n\n\n\n\n This page was last edited on 30 March 2023, at 22:55\xa0(UTC).\nText is available under the Creative Commons Attribution-ShareAlike License 3.0;\nadditional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.\n\n\nPrivacy policy\nAbout Wikipedia\nDisclaimers\nContact Wikipedia\nMobile view\nDevelopers\nStatistics\nCookie statement\n\n\n\n\n\n\n\n\n\n\n\n\n'

```
[77]: quantom_physics = quantom_physics.replace('\n', ' ')
      quantom_physics = quantom_physics.replace('\t', ' ')
      print(quantom_physics)
```

Atomic physics - WikipediaJump to contentMain menuMain menumove to sidebarhide Navigation Main pageContentsCurrent eventsRandom articleAbout WikipediaContact usDonate Contribute HelpLearn to editCommunity portalRecent changesUpload fileLanguagesLanguage links are at the top of the page across from the title.SearchCreate accountLog inPersonal tools Create account Log in Pages for logged out editors learn moreContributionsTalkContentsmove to sidebarhide(Top)1Isolated atoms2Electronic configuration3History and developments4Significant atomic physicists5See also6Bibliography7References8External links Toggle the table of contents Toggle the table of contentsAtomic physics74 languagesAfrikaansAlemannisch AsturianuAzərbaycancaBosanskiCatalàČeštinaDanskDeutschEestiΕλληνικάEspañolEuskaraFrançaisFurlanGaeilgeGalego HrvatskiBahasa IndonesiaItalianoKiswahiliLatviešuLëtzebuergeschLietuviųMagyarBahasa MelayuNederlandsNordfriiskNorsk bokmålNorsk nynorskOʻzbekcha / PlattdüütschPolskiPortuguêsRomânăSeelterskSimple EnglishSlovenčina / srpskiSrpskohrvatski / SuomiSvenskaTagalog / tatarçaTürkçeTingitVitazeuwEdit linksArticleTalkEnglishReadEditView historyToolsToolsmove to sidebarhide Actions ReadEditView history General What links hereRelated changesUpload fileSpecial pagesPermanent linkPage informationCite this pageWikidata item Print/export Download as PDFPrintable version In other projects Wikimedia CommonsFrom Wikipedia, the free encyclopediaField of physicsFor the classical view of the atom which developed into atomic physics, see atomic theory.This article includes a list of general references, but it lacks sufficient corresponding inline citations. Please help to improve this article by introducing more precise citations. (September 2015) (Learn how and when to remove this template message)Modern physics
$$\hat{H}|\psi_n(t)\rangle=i\hbar\frac{\partial}{\partial t}|\psi_n(t)\rangle$$
$$G_{\mu\nu}+\Lambda g_{\mu\nu}=\kappa T_{\mu\nu}$$
$$G_{\mu\nu}+\Lambda g_{\mu\nu}=\kappa T_{\mu\nu}$$
Schrödinger and Einstein field equationsFoundersMax Planck · Albert Einstein · Niels Bohr · Max Born · Werner Heisenberg · Erwin Schrödinger · Pascual Jordan · Wolfgang Pauli · Paul Dirac · Ernest Rutherford · Louis de Broglie · Satyendra Nath BoseConceptsTopology · Space · Time · Energy · Matter · Work Randomness · Information · Entropy · Mind Light · Particle · WaveBranchesApplied · Experimental · Theoretical Mathematical · Philosophy of physics Quantum mechanics (Quantum field theory · Quantum information · Quantum computation) Electromagnetism · Weak interaction · Electroweak interaction Strong interaction Atomic · Particle · Nuclear Atomic, molecular, and optical Condensed matter · Statistical Complex systems · Non-linear dynamics · Biophysics Neurophysics Plasma physics Special relativity · General relativity Astrophysics · Cosmology Theories of gravitation Quantum gravity · Theory of everythingScientistsWitten · Röntgen · Becquerel · Lorentz · Planck · Curie · Wien · Skłodowska-Curie · Sommerfeld · Rutherford · Soddy · Onnes · Einstein · Wilczek · Born · Weyl · Bohr · Kramers · Schrödinger · de Broglie · Laue · Bose · Compton · Pauli · Walton · Fermi · van der Waals · Heisenberg · Dyson · Zeeman · Moseley · Hilbert · Gödel · Jordan ·

Dirac · Wigner · Hawking · P. W. Anderson · Lemaître · Thomson · Poincaré ·
 Wheeler · Penrose · Millikan · Nambu · von Neumann · Higgs · Hahn · Feynman ·
 Yang · Lee · Lenard · Salam · 't Hooft · Veltman · Bell · Gell-Mann · J. J.
 Thomson · Raman · Bragg · Bardeen · Shockley · Chadwick · Lawrence ·
 Zeilinger · Goudsmit · Uhlenbeck

Categories Modern physics

Atomic physics is the field of physics that studies atoms as an isolated system of electrons and an atomic nucleus. Atomic physics typically refers to the study of atomic structure and the interaction between atoms.[1] It is primarily concerned with the way in which electrons are arranged around the nucleus and the processes by which these arrangements change. This comprises ions, neutral atoms and, unless otherwise stated, it can be assumed that the term atom includes ions. The term atomic physics can be associated with nuclear power and nuclear weapons, due to the synonymous use of atomic and nuclear in standard English. Physicists distinguish between atomic physics—which deals with the atom as a system consisting of a nucleus and electrons—and nuclear physics, which studies nuclear reactions and special properties of atomic nuclei. As with many scientific fields, strict delineation can be highly contrived and atomic physics is often considered in the wider context of atomic, molecular, and optical physics. Physics research groups are usually so classified.

Isolated atoms

Atomic physics primarily considers atoms in isolation. Atomic models will consist of a single nucleus that may be surrounded by one or more bound electrons. It is not concerned with the formation of molecules (although much of the physics is identical), nor does it examine atoms in a solid state as condensed matter. It is concerned with processes such as ionization and excitation by photons or collisions with atomic particles. While modelling atoms in isolation may not seem realistic, if one considers atoms in a gas or plasma then the time-scales for atom-atom interactions are huge in comparison to the atomic processes that are generally considered. This means that the individual atoms can be treated as if each were in isolation, as the vast majority of the time they are. By this consideration, atomic physics provides the underlying theory in plasma physics and atmospheric physics, even though both deal with very large numbers of atoms.

Electronic configuration

Electrons form notional shells around the nucleus. These are normally in a ground state but can be excited by the absorption of energy from light (photons), magnetic fields, or interaction with a colliding particle (typically ions or other electrons). In the Bohr model, the transition of an electron with $n=3$ to the shell $n=2$ is shown, where a photon is emitted. An electron from shell ($n=2$) must have been removed beforehand by ionization. Electrons that populate a shell are said to be in a bound state. The energy necessary to remove an electron from its shell (taking it to infinity) is called the binding energy. Any quantity of energy absorbed by the electron in excess of this amount is converted to kinetic energy according to the conservation of energy. The atom is said to have undergone the process of ionization. If the electron absorbs a quantity of energy less than the binding energy, it will be transferred to an excited state. After a certain time, the electron in an excited state will "jump" (undergo a transition) to a lower state. In a neutral atom, the system will emit a photon of the difference in energy, since energy is conserved. If an inner electron has absorbed more than the binding energy (so that the atom ionizes), then a more outer electron may

ISBN 978-0-387-16649-0. References[edit]^ Demtröder, W. (2006). Atoms, molecules and photons : an introduction to atomic-, molecular-, and quantum-physics. Berlin: Springer. ISBN 978-3-540-32346-4. OCLC 262692011. External links[edit] Wikimedia Commons has media related to Atomic physics. MIT-Harvard Center for Ultracold Atoms Stanford QFARM Initiative for Quantum Science & Engineering Joint Quantum Institute at University of Maryland and NIST Atomic Physics on the Internet JILA (Atomic Physics) ORNL Physics Division vte Branches of physics Divisions Pure Applied Engineering Approaches Experimental Theoretical Computational Classical Classical mechanics Newtonian Analytical Celestial Continuum Acoustics Classical electromagnetism Classical optics Ray Wave Thermodynamics Statistical Non-equilibrium Modern Relativistic mechanics Special General Nuclear physics Quantum mechanics Particle physics Atomic, molecular, and optical physics Atomic Molecular Modern optics Condensed matter physics Interdisciplinary Astrophysics Atmospheric physics Biophysics Chemical physics Geophysics Materials science Mathematical physics Medical physics Ocean physics Quantum information science Related History of physics Nobel Prize in Physics Physics education Timeline of physics discoveries Authority control National Germany Japan Czech Republic Other Encyclopedia of Modern Ukraine Retrieved from "https://en.wikipedia.org/w/index.php?title=Atomic_physics&oldid=1147433735" Categories: Atomic physics Atomic, molecular, and optical physics Hidden categories: Articles with short description Short description is different from Wikidata Articles lacking in-text citations from September 2015 All articles lacking in-text citations Articles containing Sanskrit-language text Lang and lang-xx code promoted to ISO 639-1 Articles containing German-language text Articles containing Italian-language text Articles containing Russian-language text Articles containing Macedonian-language text Commons category link from Wikidata Articles with GND identifiers Articles with NDL identifiers Articles with NKC identifiers Articles with EMU identifiers This page was last edited on 30 March 2023, at 22:55 (UTC). Text is available under the Creative Commons Attribution-ShareAlike License 3.0; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization. Privacy policy About Wikipedia Disclaimers Contact Wikipedia Mobile view Developers Statistics Cookie statement

```
[78]: from nltk.tokenize import sent_tokenize
      from nltk.tokenize import *
```

Removing the line space

```
[79]: my_token = word_tokenize(quantom_physics)
```

```
[80]: for i in my_token:
      print(i)
```

Atomic
physics
-

WikipediaJump
to
contentMain
menuMain
menumove
to
sidebarhide
Navigation
Main
pageContentsCurrent
eventsRandom
articleAbout
WikipediaContact
usDonate
Contribute
HelpLearn
to
editCommunity
portalRecent
changesUpload
fileLanguagesLanguage
links
are
at
the
top
of
the
page
across
from
the
title.SearchCreate
accountLog
inPersonal
tools
Create
account
Log
in
Pages
for
logged
out
editors
learn
moreContributionsTalkContentsmove
to

sidebarhide
 (
 Top
)
 1Isolated
 atoms2Electronic
 configuration3History
 and
 developments4Significant
 atomic
 physicists5See
 also6Bibliography7References8External
 links
 Toggle
 the
 table
 of
 contents
 Toggle
 the
 table
 of
 contentsAtomic
 physics74
 languagesAfrikaansAlemannischAsturianuAzərbaycanca
 BosanskiCatalàČeštinaDanskDeutschEestiΕλληνικάEspañolEuskaraFrançais
 sFurlanGaeilgeGalegoHrvatskiBahasa
 IndonesiaItalianoKiswahiliLatviešuLëtzebuergesch
 LietuviųMagyarBahasa
 MelayuNederlandsNordfriiskNorsk
 bokmålNorsk
 nynorskOzbekcha
 /
 PlattdüütschPolskiPortuguêsRomânăSeelterskSimple
 EnglishSlovenčina
 /
 srpskiSrpskohrvatski
 /
 SuomiSvenskaTagalog
 /
 tatarçaTürkçeTing
 VitZeêuwsEdit
 linksArticleTalkEnglishReadEditView
 historyToolsToolsmove
 to
 sidebarhide
 Actions
 ReadEditView

to
the
Terms
of
Use
and
Privacy
Policy
.
Wikipedia®
is
a
registered
trademark
of
the
Wikimedia
Foundation
,
Inc.
,
a
non-profit
organization.Privacy
policyAbout
WikipediaDisclaimersContact
WikipediaMobile
viewDevelopersStatisticsCookie
statement

PART 4 | Create a corpus from Text

```
[83]: print("Text reader")
      from nltk.corpus.reader import PlaintextCorpusReader
      corpus = PlaintextCorpusReader('C:\\Users\\ayman\\OneDrive - Istanbul Bilgi_
      ↪Universitesi\\Coding-New\\Coding\\WORKSPACE\\PYTHON\\machine_
      ↪translation\\lab5','cars.text')
```

Text reader

```
[84]: corpus.words()
```

```
[84]: ['Ford', 'Model', 'T', 'Volkswagen', 'Beetle', ...]
```

```
[85]: print(corpus.fileids())
```

```
['cars.text']
```

```
[86]: print(len(corpus.words()))
```

127

```
[88]: for i in corpus.words():  
      print(i)
```

Ford
Model
T
Volkswagen
Beetle
Porsche
911
Mini
Cooper
Jaguar
E
-
Type
Chevrolet
Corvette
Ferrari
250
GT0
Lamborghini
Miura
Ford
Mustang
Dodge
Charger
Pontiac
GT0
Chevrolet
Camaro
Plymouth
Barracuda
Chevrolet
Bel
Air
Cadillac
Eldorado
Rolls
-
Royce
Silver
Ghost
Aston
Martin
DB5

Mercedes
-
Benz
300SL
BMW
507
Fiat
500
Alfa
Romeo
Spider
Toyota
2000GT
Nissan
Skyline
GT
-
R
Honda
NSX
Mazda
RX
-
7
Dodge
Viper
Plymouth
Superbird
DeLorean
DMC
-
12
Audi
Quattro
Ferrari
Testarossa
Lamborghini
Countach
Ford
GT40
Porsche
959
BMW
M1
McLaren
F1
Honda
Civic

Type
R
Subaru
Impreza
WRX
STI
Mitsubishi
Lancer
Evolution
Ford
RS200
Audi
RS2
Avant
Toyota
Supra
Lexus
LFA
Porsche
918
Spyder
McLaren
P1
Lamborghini
Aventador
Bugatti
Veyron
Ferrari
LaFerrari
Koenigsegg
Agera
Pagani
Huayra
Tesla
Model
S

PART 5 | IBM MODEL

```
[89]: print ("test IBM model 1")  
      from nltk.translate import AlignedSent, Alignment, IBMModel1  
      print (" packages IBM model 1 imported")
```

```
test IBM model 1  
packages IBM model 1 imported
```

```
[90]: bitext= []  
      bitext.append(AlignedSent(['klein', 'ist', 'das', 'haus'],
```



```

['the', 'hauser', 'is', 'small'])
bitext.append(AlignedSent(['das', 'haus', 'ist', 'ja', 'groß'],
['the', 'house', 'is', 'big']))
bitext.append(AlignedSent(['das', 'buch', 'ist', 'ja', 'klein'],
['the', 'book', 'is', 'small']))
bitext.append(AlignedSent(['das', 'haus'], ['the', 'house']))
bitext.append(AlignedSent(['das', 'buch'], ['the', 'book']))
bitext.append(AlignedSent(['ein', 'buch'], ['a', 'book']))
print(bitext)
myIBM = IBMModel1(bitext, 5)

```

```

[AlignedSent(['klein', 'ist', 'das', 'haus'], ['the', 'hauser', 'is', 'small'],
Alignment([])), AlignedSent(['das', 'haus', 'ist', 'ja', 'groß'], ['the',
'house', 'is', 'big'], Alignment([])), AlignedSent(['das', 'buch', 'ist', 'ja',
'klein'], ['the', 'book', 'is', 'small'], Alignment([])), AlignedSent(['das',
'haus'], ['the', 'house'], Alignment([])), AlignedSent(['das', 'buch'], ['the',
'book'], Alignment([])), AlignedSent(['ein', 'buch'], ['a', 'book'],
Alignment([]))]

```

```

[91]: print("translate")
print(myIBM.translation_table['buch']['book'])
print(myIBM.translation_table['das']['the'])

```

```

translate
0.8884662872538488
0.6012907706192668

```

```

[92]: print("test -one-by-one")
test_sentence= bitext[0]
print(test_sentence.words)

```

```

test -one-by-one
['klein', 'ist', 'das', 'haus']

```

```

[93]: print(test_sentence.mots)
print(test_sentence.alignment)

```

```

['the', 'hauser', 'is', 'small']
0-3 1-2 2-0 3-1

```

```

[94]: print(" check the 3rd sentence")
test_sentence= bitext[3]
print(test_sentence.words)
print(test_sentence.mots)
print(test_sentence.alignment)

```

```

check the 3rd sentence
['das', 'haus']
['the', 'house']
0-0 1-1

```

```
[95]: print(" check the 2nd sentence")
      test_sentence= bitext[2]
      print(test_sentence.words)
      print(test_sentence.mots)
      print(test_sentence.alignment)

      check the 2nd sentence
      ['das', 'buch', 'ist', 'ja', 'klein']
      ['the', 'book', 'is', 'small']
      0-0 1-1 2-2 3-2 4-3
```