READ ME

**Assumptions:**

1.The zip file is extracted and gave the paths accordingly.

2.All the special characters are replaced by space and ‘s with null string.

3. Tokenization is performed after eliminating all the special characters and then tokens are split by space.

4.Named Entity recognition has been found using string matching but it can further be improved a lot by adding heuristics such as capitalization of words, training a machine learning algorithm to basically account for the context in which the word has appeared (considering POS tagging and other features).

5.We could also us e class labels to classify the recognized NER’s.

**Limitations:**

1.Identified sentences using Capitalization after special symbols like(:,?,!,.,,,).

2.Did not use semantic parameters like dependent clauses to identify the sentences.

**Steps to execute:**

1.Parameters are passed to main method.Path of NER.txt is given as arg[1] and path of nlp\_data which has all data files in args[0].

2.MainProcess is the file where the entire execution starts.

**Output:**

pool-1-thread-1 fileName C:\Users\Sahithi\Documents\programming\_test\NLP\_test\nlp\_data\d01.txt

----------------------------------------------------

Program 1: Identified SentencesEuclid's Elements has been referred to as the most successful and influential textbook ever written.

Program 1: Identified Tokens{been=1, textbook=1, influential=1, Euclid=1, Elements=1, the=1, most=1, ever=1, as=1, and=1, has=1, to=1, written=1, referred=1, successful=1}

Program 2: Identified Named Entities[Euclid, Elements]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesFirst set in type in Venice in 1482, it is one of the very earliest mathematical works to be printed after the invention of the printing press and was estimated by Carl Benjamin Boyer to be second only to the Bible in the number of editions published, with the number reaching well over one thousand.

Program 1: Identified Tokens{very=1, works=1, be=2, Carl Benjamin Boyer=1, Bible=1, 1482=1, reaching=1, type=1, second=1, number=2, and=1, of=3, by=1, only=1, First=1, after=1, press=1, over=1, set=1, in=4, one=2, was=1, estimated=1, is=1, it=1, published=1, earliest=1, printing=1, Venice=1, the=6, printed=1, with=1, editions=1, well=1, to=3, invention=1, thousand=1, mathematical=1}

Program 2: Identified Named Entities[Venice, Carl Benjamin Boyer, Bible]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesFor centuries, knowledge of at least part of the Elements was required of all students.

Program 1: Identified Tokens{all=1, centuries=1, part=1, For=1, least=1, was=1, students=1, Elements=1, required=1, the=1, at=1, of=3, knowledge=1}

Program 2: Identified Named Entities[Elements]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesNot until the 20th century, by which time its content was universally taught through other school textbooks, did it cease to be considered something all educated people had read.

Program 1: Identified Tokens{through=1, other=1, be=1, considered=1, universally=1, had=1, textbooks=1, content=1, school=1, by=1, which=1, all=1, read=1, 20th=1, was=1, its=1, it=1, people=1, something=1, the=1, Not=1, century=1, educated=1, cease=1, taught=1, until=1, time=1, to=1, did=1}

Program 2: Identified Named Entities[]

----------------------------------------------------

pool-1-thread-1 (End)

pool-1-thread-1 fileName C:\Users\Sahithi\Documents\programming\_test\NLP\_test\nlp\_data\d02.txt

----------------------------------------------------

Program 1: Identified SentencesThe Broydenâ€“Fletcherâ€“Goldfarbâ€“Shanno (BFGS) algorithm is an iterative method for solving unconstrained nonlinear optimization problems.

Program 1: Identified Tokens{solving=1, method=1, for=1, iterative=1, is=1, nonlinear=1, an=1, BFGS=1, unconstrained=1, The=1, optimization=1, Broydenâ€“Fletcherâ€“Goldfarbâ€“Shanno=1, algorithm=1, problems=1}

Program 2: Identified Named Entities[Broydenâ€“Fletcherâ€“Goldfarbâ€“Shanno, BFGS]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesThe BFGS method approximates Newton's method, a class of hill-climbing optimization techniques that seeks a stationary point of a (preferably twice continuously differentiable) function.

Program 1: Identified Tokens{a=3, method=2, preferably=1, continuously=1, differentiable=1, point=1, BFGS=1, The=1, that=1, Newton=1, twice=1, optimization=1, stationary=1, techniques=1, of=2, function=1, class=1, approximates=1, hill-climbing=1, seeks=1}

Program 2: Identified Named Entities[Newton, BFGS]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesFor such problems, a necessary condition for optimality is that the gradient be zero.

Program 1: Identified Tokens{a=1, be=1, necessary=1, optimality=1, gradient=1, For=1, for=1, is=1, the=1, zero=1, such=1, that=1, condition=1, problems=1}

Program 2: Identified Named Entities[]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesNewton's method and the BFGS methods are not guaranteed to converge unless the function has a quadratic Taylor expansion near an optimum.

Program 1: Identified Tokens{a=1, method=1, Taylor=1, methods=1, optimum=1, guaranteed=1, quadratic=1, an=1, BFGS=1, expansion=1, the=2, Newton=1, not=1, converge=1, are=1, unless=1, and=1, function=1, to=1, has=1, near=1}

Program 2: Identified Named Entities[Newton, BFGS]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesThese methods use both the first and second derivatives of the function.

Program 1: Identified Tokens{the=2, derivatives=1, These=1, methods=1, use=1, and=1, of=1, function=1, first=1, both=1, second=1}

Program 2: Identified Named Entities[]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesHowever, BFGS has proven to have good performance even for non-smooth optimizations.

Program 1: Identified Tokens{performance=1, even=1, have=1, for=1, However=1, has=1, proven=1, to=1, non-smooth=1, good=1, BFGS=1, optimizations=1}

Program 2: Identified Named Entities[BFGS]

----------------------------------------------------

pool-1-thread-1 (End)

pool-1-thread-1 fileName C:\Users\Sahithi\Documents\programming\_test\NLP\_test\nlp\_data\d03.txt

----------------------------------------------------

Program 1: Identified SentencesMontgomery Castle is a stone masonry castle looking over the town of Montgomery in Powys, Wales.

Program 1: Identified Tokens{over=1, a=1, castle=1, town=1, in=1, Montgomery=1, Wales=1, Powys=1, is=1, stone=1, Montgomery Castle=1, masonry=1, the=1, of=1, looking=1}

Program 2: Identified Named Entities[Montgomery Castle, Powys, Wales]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesIt is one of many Norman castles on the border between Wales and England.

Program 1: Identified Tokens{border=1, Norman=1, castles=1, one=1, Wales=1, England=1, is=1, It=1, many=1, the=1, and=1, of=1, between=1, on=1}

Program 2: Identified Named Entities[Wales, England]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesThe original castle was built at the order of Roger de Montgomery, earl of Shrewsbury sometime between 1071 and 1074.

Program 1: Identified Tokens{sometime=1, original=1, castle=1, built=1, was=1, Roger de Montgomery=1, 1074=1, 1071=1, The=1, the=1, at=1, Shrewsbury=1, and=1, of=2, earl=1, between=1, order=1}

Program 2: Identified Named Entities[Roger de Montgomery, Shrewsbury]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesOn the rebellion of his son Robert of Belleme in 1102, the castle was given to Baldwin de Boulers.

Program 1: Identified Tokens{given=1, de=1, rebellion=1, castle=1, in=1, Baldwin=1, was=1, Boulers=1, the=2, his=1, son=1, 1102=1, of=1, to=1, On=1, Robert of Belleme=1}

Program 2: Identified Named Entities[Robert of Belleme]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesIn 1267 Montgomery was the meeting place for treaty negotiations, where King Henry III granted Llywelyn ap Gruffudd the title of Prince of Wales.

Program 1: Identified Tokens{In=1, Montgomery=1, was=1, for=1, negotiations=1, title=1, granted=1, Prince of Wales=1, the=2, 1267=1, treaty=1, of=1, Llywelyn ap Gruffudd=1, where=1, place=1, King Henry III=1, meeting=1}

Program 2: Identified Named Entities[Wales, King Henry III, Llywelyn ap Gruffudd, Prince of Wales]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesAfter 1295 and the final Welsh War of the thirteenth century the castle became more of a military backwater and prison than a front line fortress.

Program 1: Identified Tokens{became=1, a=2, castle=1, more=1, line=1, War=1, 1295=1, Welsh=1, thirteenth=1, backwater=1, the=3, century=1, fortress=1, and=2, of=2, than=1, After=1, final=1, military=1, front=1, prison=1}

Program 2: Identified Named Entities[]

----------------------------------------------------

pool-1-thread-1 (End)

pool-1-thread-1 fileName C:\Users\Sahithi\Documents\programming\_test\NLP\_test\nlp\_data\d04.txt

----------------------------------------------------

Program 1: Identified SentencesThe Antikythera mechanism is an ancient analog computer designed to predict astronomical positions and eclipses for calendrical and astrological purposes, as well as the cycles of Olympic Games.Found housed in a 340 Ã— 180 Ã— 90 mm wooden box, the device is a complex clockwork mechanism composed of at least 30 meshing bronze gears.

Program 1: Identified Tokens{astronomical=1, for=1, least=1, cycles=1, box=1, clockwork=1, calendrical=1, housed=1, The=1, computer=1, composed=1, and=2, of=2, 90=1, complex=1, predict=1, mechanism=2, designed=1, 30=1, gears=1, mm=1, a=2, in=1, 180=1, is=2, positions=1, eclipses=1, GamesFound=1, an=1, ancient=1, Olympic=1, 340=1, bronze=1, the=2, analog=1, Ã—=2, as=2, at=1, meshing=1, Antikythera=1, purposes=1, wooden=1, well=1, to=1, astrological=1, device=1}

Program 2: Identified Named Entities[Antikythera]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesIts remains were found as 82 separate fragments, of which only seven contain any gears or significant inscriptions.

Program 1: Identified Tokens{which=1, contain=1, or=1, separate=1, inscriptions=1, Its=1, seven=1, remains=1, significant=1, any=1, as=1, found=1, were=1, of=1, only=1, 82=1, fragments=1, gears=1}

Program 2: Identified Named Entities[]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesThe largest gear is approximately 140 mm in diameter.The artifact was recovered in 1901 from the Antikythera shipwreck off the Greek island of Antikythera.

Program 1: Identified Tokens{mm=1, largest=1, in=2, island=1, was=1, is=1, diameterThe=1, 140=1, off=1, The=1, artifact=1, the=2, shipwreck=1, recovered=1, approximately=1, Antikythera=2, of=1, from=1, Greek=1, 1901=1, gear=1}

Program 2: Identified Named Entities[Antikythera]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesBelieved to have been designed and constructed by Greek scientists, the instrument has been dated either between 150 to 100 BC or, according to a more recent view, at 205 BC.

Program 1: Identified Tokens{BC=2, been=2, according=1, instrument=1, 150=1, either=1, view=1, Believed=1, and=1, by=1, have=1, dated=1, has=1, designed=1, recent=1, between=1, a=1, or=1, constructed=1, more=1, the=1, 100=1, at=1, scientists=1, 205=1, to=3, Greek=1}

Program 2: Identified Named Entities[]

----------------------------------------------------

pool-1-thread-1 (End)

pool-1-thread-1 fileName C:\Users\Sahithi\Documents\programming\_test\NLP\_test\nlp\_data\d05.txt

----------------------------------------------------

Program 1: Identified SentencesHorseshoe crabs are marine arthropods that live primarily in and around shallow ocean waters on soft sandy or muddy bottoms.

Program 1: Identified Tokens{waters=1, Horseshoe=1, or=1, in=1, arthropods=1, muddy=1, crabs=1, around=1, bottoms=1, marine=1, primarily=1, that=1, ocean=1, are=1, and=1, sandy=1, shallow=1, live=1, on=1, soft=1}

Program 2: Identified Named Entities[]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesThey occasionally come onto shore to mate.

Program 1: Identified Tokens{mate=1, occasionally=1, shore=1, come=1, to=1, onto=1, They=1}

Program 2: Identified Named Entities[]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesThey are commonly used as bait and in fertilizer.

Program 1: Identified Tokens{as=1, commonly=1, bait=1, are=1, in=1, and=1, used=1, fertilizer=1, They=1}

Program 2: Identified Named Entities[]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesIn recent years, a decline in the population has occurred as a consequence of coastal habitat destruction in Japan and overharvesting along the east coast of North America.

Program 1: Identified Tokens{a=2, consequence=1, habitat=1, occurred=1, In=1, in=2, Japan=1, overharvesting=1, North America=1, years=1, population=1, the=2, coast=1, as=1, east=1, along=1, and=1, of=2, decline=1, has=1, recent=1, coastal=1, destruction=1}

Program 2: Identified Named Entities[Japan, North America]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesBecause of their origin 450 million years ago (Mya), horseshoe crabs are considered living fossils.

Program 1: Identified Tokens{living=1, considered=1, origin=1, ago=1, fossils=1, their=1, crabs=1, years=1, 450=1, horseshoe=1, Because=1, million=1, Mya=1, are=1, of=1}

Program 2: Identified Named Entities[]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesHorseshoe crabs resemble crustaceans, but belong to a separate subphylum, and are closely related to arachnids.

Program 1: Identified Tokens{but=1, a=1, Horseshoe=1, belong=1, closely=1, crustaceans=1, separate=1, resemble=1, crabs=1, arachnids=1, related=1, are=1, and=1, subphylum=1, to=2}

Program 2: Identified Named Entities[]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesThe earliest horseshoe crab fossils are found in strata from the late Ordovician period, roughly 450 Mya.

Program 1: Identified Tokens{period=1, in=1, fossils=1, strata=1, Ordovician=1, earliest=1, 450=1, The=1, horseshoe=1, the=1, found=1, late=1, Mya=1, are=1, roughly=1, from=1, crab=1}

Program 2: Identified Named Entities[]

----------------------------------------------------

pool-1-thread-1 (End)

pool-1-thread-1 fileName C:\Users\Sahithi\Documents\programming\_test\NLP\_test\nlp\_data\d06.txt

----------------------------------------------------

Program 1: Identified SentencesA differentiable manifold is a type of manifold that is locally similar enough to a linear space to allow one to do calculus.

Program 1: Identified Tokens{allow=1, A=1, a=2, similar=1, linear=1, one=1, is=2, enough=1, do=1, differentiable=1, type=1, calculus=1, space=1, that=1, locally=1, of=1, manifold=2, to=3}

Program 2: Identified Named Entities[]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesAny manifold can be described by a collection of charts, also known as an atlas.

Program 1: Identified Tokens{a=1, charts=1, be=1, atlas=1, also=1, collection=1, Any=1, an=1, can=1, as=1, known=1, by=1, of=1, manifold=1, described=1}

Program 2: Identified Named Entities[]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesOne may then apply ideas from calculus while working within the individual charts, since each chart lies within a linear space to which the usual rules of calculus apply.

Program 1: Identified Tokens{charts=1, One=1, rules=1, calculus=2, while=1, space=1, of=1, working=1, lies=1, from=1, which=1, a=1, linear=1, may=1, individual=1, usual=1, apply=2, within=2, ideas=1, then=1, each=1, the=2, to=1, chart=1, since=1}

Program 2: Identified Named Entities[]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesIf the charts are suitably compatible (namely, the transition from one chart to another is differentiable), then computations done in one chart are valid in any other differentiable chart.

Program 1: Identified Tokens{charts=1, other=1, computations=1, in=2, suitably=1, namely=1, one=2, another=1, is=1, then=1, differentiable=2, done=1, any=1, transition=1, the=2, valid=1, compatible=1, are=2, from=1, to=1, If=1, chart=3}

Program 2: Identified Named Entities[]

----------------------------------------------------

pool-1-thread-1 (End)

pool-1-thread-1 fileName C:\Users\Sahithi\Documents\programming\_test\NLP\_test\nlp\_data\d07.txt

----------------------------------------------------

Program 1: Identified SentencesSun Microsystems, Inc. was a company that sold computers, computer components, computer software, and information technology services and that created the Java programming language and the Network File System (NFS).

Program 1: Identified Tokens{components=1, software=1, Sun Microsystems=1, language=1, System=1, that=2, computer=2, and=3, Network=1, company=1, NFS=1, computers=1, programming=1, a=1, sold=1, Java=1, created=1, was=1, technology=1, services=1, the=2, information=1, File=1, Inc=1}

Program 2: Identified Named Entities[Sun Microsystems]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesSun significantly evolved several key computing technologies, among them Unix, RISC processors, thin client computing, and virtualized computing.

Program 1: Identified Tokens{Unix=1, RISC=1, evolved=1, several=1, them=1, processors=1, Sun=1, thin=1, technologies=1, computing=3, virtualized=1, among=1, and=1, client=1, significantly=1, key=1}

Program 2: Identified Named Entities[Sun]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesSun was founded on February 24, 1982.On January 27, 2010, Sun was acquired by Oracle Corporation for US $7.4 billion.

Program 1: Identified Tokens{24=1, 27=1, was=2, founded=1, for=1, 2010=1, Sun=2, billion=1, by=1, 1982On=1, February=1, January=1, 74=1, acquired=1, US=1, on=1, Oracle Corporation=1}

Program 2: Identified Named Entities[Sun, Oracle Corporation]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesThe following month, Sun was merged with Oracle USA, Inc. to become Oracle America, Inc.

Program 1: Identified Tokens{USA=1, was=1, merged=1, Sun=1, The=1, with=1, month=1, following=1, America=1, to=1, Oracle=2, Inc=2, become=1}

Program 2: Identified Named Entities[Sun]

----------------------------------------------------

pool-1-thread-1 (End)

pool-1-thread-1 fileName C:\Users\Sahithi\Documents\programming\_test\NLP\_test\nlp\_data\d08.txt

----------------------------------------------------

Program 1: Identified SentencesApollo 11 was the spaceflight that landed the first humans on the Moon, Americans Neil Armstrong and Buzz Aldrin, on July 20, 1969.

Program 1: Identified Tokens{Neil Armstrong=1, Moon=1, Americans=1, was=1, Buzz Aldrin=1, the=3, spaceflight=1, that=1, and=1, Apollo 11=1, July=1, 1969=1, humans=1, first=1, 20=1, landed=1, on=2}

Program 2: Identified Named Entities[Apollo 11, Neil Armstrong, Buzz Aldrin, Moon]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesArmstrong became the first to step onto the lunar surface six hours later.

Program 1: Identified Tokens{the=2, Armstrong=1, became=1, hours=1, six=1, later=1, surface=1, step=1, to=1, first=1, onto=1, lunar=1}

Program 2: Identified Named Entities[]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesArmstrong spent about two and a half hours outside the spacecraft, Aldrin slightly less, and together they collected 47.5 pounds (21.5 kg) of lunar material for return to Earth.

Program 1: Identified Tokens{slightly=1, half=1, about=1, for=1, collected=1, two=1, Earth=1, Armstrong=1, spacecraft=1, 475=1, and=2, pounds=1, 215=1, of=1, kg=1, together=1, they=1, lunar=1, a=1, hours=1, spent=1, less=1, the=1, Aldrin=1, material=1, outside=1, to=1, return=1}

Program 2: Identified Named Entities[Earth]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesThe third member of the mission, Michael Collins, piloted the command spacecraft alone in lunar orbit until Armstrong and Aldrin returned to it just under a day later for the trip back to Earth.

Program 1: Identified Tokens{for=1, back=1, Michael Collins=1, The=1, Earth=1, spacecraft=1, Armstrong=1, alone=1, later=1, trip=1, and=1, of=1, member=1, piloted=1, orbit=1, just=1, day=1, lunar=1, a=1, in=1, it=1, command=1, the=3, mission=1, Aldrin=1, third=1, until=1, to=2, returned=1, under=1}

Program 2: Identified Named Entities[Michael Collins, Earth]

----------------------------------------------------

pool-1-thread-1 (End)

pool-1-thread-1 fileName C:\Users\Sahithi\Documents\programming\_test\NLP\_test\nlp\_data\d09.txt

----------------------------------------------------

Program 1: Identified SentencesThe Apollo spacecraft had three parts: a Command Module with a cabin for the three astronauts, and the only part that landed back on Earth; a Service Module, which supported the Command Module with propulsion, electrical power, oxygen, and water; and a Lunar Module for landing on the Moon.

Program 1: Identified Tokens{astronauts=1, Moon=1, part=1, for=2, cabin=1, back=1, had=1, three=2, The=1, Earth=1, spacecraft=1, that=1, and=3, only=1, power=1, oxygen=1, landed=1, on=2, which=1, a=4, Apollo=1, Service=1, propulsion=1, electrical=1, water=1, the=4, parts:=1, with=2, Lunar=1, landing=1, Command=2, Module=4, supported=1}

Program 2: Identified Named Entities[Moon, Earth]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesAfter being sent toward the Moon by the Saturn V's upper stage, the astronauts separated the spacecraft from it and traveled for three days until they entered into lunar orbit.

Program 1: Identified Tokens{Saturn=1, astronauts=1, Moon=1, upper=1, for=1, three=1, spacecraft=1, into=1, V=1, and=1, by=1, from=1, orbit=1, they=1, lunar=1, separated=1, being=1, it=1, entered=1, sent=1, the=4, toward=1, stage=1, After=1, days=1, traveled=1, until=1}

Program 2: Identified Named Entities[Moon]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesArmstrong and Aldrin then moved into the Lunar Module and landed in the Sea of Tranquility.

Program 1: Identified Tokens{Sea of Tranquility=1, in=1, moved=1, then=1, the=2, Armstrong=1, into=1, Aldrin=1, Lunar=1, and=2, Module=1, landed=1}

Program 2: Identified Named Entities[Sea of Tranquility]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesThey stayed a total of about 21.5 hours on the lunar surface.

Program 1: Identified Tokens{the=1, a=1, total=1, hours=1, surface=1, stayed=1, 215=1, of=1, about=1, lunar=1, They=1, on=1}

Program 2: Identified Named Entities[]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesAfter lifting off in the upper part of the Lunar Module and rejoining Collins in the Command Module, they returned to Earth and landed in the Pacific Ocean on July 24.

Program 1: Identified Tokens{24=1, in=3, upper=1, part=1, rejoining=1, off=1, the=4, Earth=1, Collins=1, Lunar=1, Command=1, lifting=1, and=2, of=1, After=1, Pacific Ocean=1, to=1, July=1, Module=2, returned=1, they=1, landed=1, on=1}

Program 2: Identified Named Entities[Earth, Pacific Ocean]

----------------------------------------------------

pool-1-thread-1 (End)

pool-1-thread-1 fileName C:\Users\Sahithi\Documents\programming\_test\NLP\_test\nlp\_data\d10.txt

----------------------------------------------------

Program 1: Identified SentencesJames Clerk Maxwell was a Scottish mathematical physicist.

Program 1: Identified Tokens{a=1, Scottish=1, was=1, physicist=1, mathematical=1, James Clerk Maxwell=1}

Program 2: Identified Named Entities[James Clerk Maxwell]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesHis most notable achievement was to formulate the classical theory of electromagnetic radiation, bringing together for the first time electricity, magnetism, and light as manifestations of the same phenomenon.

Program 1: Identified Tokens{manifestations=1, achievement=1, formulate=1, classical=1, for=1, magnetism=1, notable=1, phenomenon=1, same=1, and=1, of=2, together=1, was=1, electricity=1, most=1, the=3, as=1, His=1, bringing=1, electromagnetic=1, radiation=1, light=1, to=1, time=1, theory=1, first=1}

Program 2: Identified Named Entities[]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesMaxwell's equations for electromagnetism have been called the "second great unification in physics" after the first one realised by Isaac Newton...

Program 1: Identified Tokens{=7, been=1, realised=1, electromagnetism=1, called=1, in=1, one=1, for=1, Maxwell=1, unification=1, great=1, second=1, equations=1, the=2, physics=1, by=1, have=1, Isaac Newton=1, after=1, first=1}

Program 2: Identified Named Entities[Newton, Isaac Newton]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesMaxwell demonstrated that electric and magnetic fields travel through space as waves moving at the speed of light.

Program 1: Identified Tokens{through=1, waves=1, electric=1, Maxwell=1, space=1, speed=1, the=1, that=1, as=1, at=1, light=1, and=1, of=1, magnetic=1, demonstrated=1, fields=1, travel=1, moving=1}

Program 2: Identified Named Entities[]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesMaxwell proposed that light is an undulation in the same medium that is the cause of electric and magnetic phenomena.

Program 1: Identified Tokens{in=1, electric=1, cause=1, Maxwell=1, is=2, medium=1, undulation=1, an=1, the=2, that=2, same=1, proposed=1, light=1, and=1, of=1, magnetic=1, phenomena=1}

Program 2: Identified Named Entities[]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesThe unification of light and electrical phenomena led to the prediction of the existence of radio waves.

Program 1: Identified Tokens{=4, waves=1, existence=1, electrical=1, led=1, unification=1, radio=1, The=1, the=2, light=1, and=1, of=3, prediction=1, to=1, phenomena=1}

Program 2: Identified Named Entities[]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesHis discoveries helped usher in the era of modern physics, laying the foundation for such fields as special relativity and quantum mechanics.

Program 1: Identified Tokens{laying=1, in=1, for=1, quantum=1, mechanics=1, foundation=1, the=2, special=1, usher=1, such=1, as=1, His=1, modern=1, discoveries=1, era=1, relativity=1, and=1, of=1, physics=1, helped=1, fields=1}

Program 2: Identified Named Entities[]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesHis contributions to the science are considered by many to be of the same magnitude as those of Isaac Newton and Albert Einstein.

Program 1: Identified Tokens{be=1, considered=1, many=1, the=2, as=1, His=1, contributions=1, same=1, are=1, and=1, science=1, by=1, of=2, Isaac Newton=1, magnitude=1, to=2, those=1, Albert Einstein=1}

Program 2: Identified Named Entities[Newton, Isaac Newton, Albert Einstein]

----------------------------------------------------

----------------------------------------------------

Program 1: Identified SentencesQED

Program 1: Identified Tokens{QED=1}

Program 2: Identified Named Entities[]

----------------------------------------------------

pool-1-thread-1 (End)