

In [1]:

```
# BBC Dataset: Athletics folder only (first line is heading/summary, rest are content)
```

In [2]:

```
#Claxton hunting first major medal

#British hurdler Sarah Claxton is confident she can win her first major medal at next month
# European Indoor Championships in Madrid.

# The 25-year-old has already smashed the British record over 60m hurdles twice this season
# setting a new mark of 7.96 seconds to win the AAAs title. "I am quite confident," said CL
# "But I take each race as it comes.
# "As long as I keep up my training but not do too much I think there is a chance of a meda
# " Claxton has won the national 60m hurdles title for the past three years but has struggl
# translate her domestic success to the international stage.
# Now, the Scotland-born athlete owns the equal fifth-fastest time in the world this year.
# And at last week's Birmingham Grand Prix, Claxton left European medal favourite Russian I
# trailing in sixth spot.

#For the first time, Claxton has only been preparing for a campaign over the hurdles - whic
# In previous seasons, the 25-year-old also contested the long jump but since moving from C
# she has re-focused her attentions. Claxton will see if her new training regime pays divid
# European Indoors which take place on 5-6 March.
```

In [3]:

```
!pip install nltk
```

```
Requirement already satisfied: nltk in c:\programdata\anaconda3\lib\site-pac
kages (3.6.5)
Requirement already satisfied: click in c:\programdata\anaconda3\lib\site-pa
ckages (from nltk) (8.0.3)
Requirement already satisfied: joblib in c:\programdata\anaconda3\lib\site-p
ackages (from nltk) (1.1.0)
Requirement already satisfied: regex>=2021.8.3 in c:\programdata\anaconda3\l
ib\site-packages (from nltk) (2021.8.3)
Requirement already satisfied: tqdm in c:\programdata\anaconda3\lib\site-pac
kages (from nltk) (4.62.3)
Requirement already satisfied: colorama in c:\programdata\anaconda3\lib\site
-packages (from click->nltk) (0.4.4)
```

In [4]:

```
import nltk
nltk.download('punkt')
```

```
[nltk_data] Downloading package punkt to
[nltk_data] C:\Users\Administrator\AppData\Roaming\nltk_data...
[nltk_data] Package punkt is already up-to-date!
```

Out[4]:

True

In [5]:

```
!pip install sumy
```

Collecting sumy

Downloading sumy-0.9.0-py2.py3-none-any.whl (87 kB)

Requirement already satisfied: nltk>=3.0.2 in c:\programdata\anaconda3\lib\site-packages (from sumy) (3.6.5)

Requirement already satisfied: requests>=2.7.0 in c:\programdata\anaconda3\lib\site-packages (from sumy) (2.26.0)

Collecting docopt<0.7,>=0.6.1

Downloading docopt-0.6.2.tar.gz (25 kB)

Collecting pycountry>=18.2.23

Downloading pycountry-22.3.5.tar.gz (10.1 MB)

Installing build dependencies: started

Installing build dependencies: finished with status 'done'

Getting requirements to build wheel: started

Getting requirements to build wheel: finished with status 'done'

Preparing wheel metadata: started

Preparing wheel metadata: finished with status 'done'

Collecting breadability>=0.1.20

Downloading breadability-0.1.20.tar.gz (32 kB)

Requirement already satisfied: chardet in c:\programdata\anaconda3\lib\site-packages (from breadability>=0.1.20->sumy) (4.0.0)

Requirement already satisfied: lxml>=2.0 in c:\programdata\anaconda3\lib\site-packages (from breadability>=0.1.20->sumy) (4.6.3)

Requirement already satisfied: click in c:\programdata\anaconda3\lib\site-packages (from nltk>=3.0.2->sumy) (8.0.3)

Requirement already satisfied: joblib in c:\programdata\anaconda3\lib\site-packages (from nltk>=3.0.2->sumy) (1.1.0)

Requirement already satisfied: regex>=2021.8.3 in c:\programdata\anaconda3\lib\site-packages (from nltk>=3.0.2->sumy) (2021.8.3)

Requirement already satisfied: tqdm in c:\programdata\anaconda3\lib\site-packages (from nltk>=3.0.2->sumy) (4.62.3)

Requirement already satisfied: setuptools in c:\programdata\anaconda3\lib\site-packages (from pycountry>=18.2.23->sumy) (58.0.4)

Requirement already satisfied: charset-normalizer~2.0.0 in c:\programdata\anaconda3\lib\site-packages (from requests>=2.7.0->sumy) (2.0.4)

Requirement already satisfied: idna<4,>=2.5 in c:\programdata\anaconda3\lib\site-packages (from requests>=2.7.0->sumy) (3.2)

Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\programdata\anaconda3\lib\site-packages (from requests>=2.7.0->sumy) (1.26.7)

Requirement already satisfied: certifi>=2017.4.17 in c:\programdata\anaconda3\lib\site-packages (from requests>=2.7.0->sumy) (2021.10.8)

Requirement already satisfied: colorama in c:\programdata\anaconda3\lib\site-packages (from click->nltk>=3.0.2->sumy) (0.4.4)

Building wheels for collected packages: breadability, docopt, pycountry

Building wheel for breadability (setup.py): started

Building wheel for breadability (setup.py): finished with status 'done'

Created wheel for breadability: filename=breadability-0.1.20-py2.py3-none-any.whl size=21711 sha256=8accd968873ab7368f634ee3f450c790797b9e07152e1dd924ce24e6c10a3d44

Stored in directory: c:\users\administrator\appdata\local\pip\cache\wheels\ba\9f\70\7795228568b81b57a8932755938da9fb1f291b0576752604aa

Building wheel for docopt (setup.py): started

Building wheel for docopt (setup.py): finished with status 'done'

Created wheel for docopt: filename=docopt-0.6.2-py2.py3-none-any.whl size=13724 sha256=72ba7154796e0987b16fc7a429aa78d0dd7f1da9c97c558507babe8fb810525d

Stored in directory: c:\users\administrator\appdata\local\pip\cache\wheels\70\4a\46\1309fc853b8d395e60bafaf1b6df7845bdd82c95fd59dd8d2b

```
Building wheel for pycountry (PEP 517): started
Building wheel for pycountry (PEP 517): finished with status 'done'
Created wheel for pycountry: filename=pycountry-22.3.5-py2.py3-none-any.whl
size=10681845 sha256=d7fcf7528888c09dfa1275c9053df0887192cb46d7faed1896c00
cbe6025da41
Stored in directory: c:\users\administrator\appdata\local\pip\cache\wheels
\47\15\92\e6dc85fcb0686c82e1edbcfdf80cfe4808c058813fed0baa8f
Successfully built breadability docopt pycountry
Installing collected packages: docopt, pycountry, breadability, sumy
Successfully installed breadability-0.1.20 docopt-0.6.2 pycountry-22.3.5 sum
y-0.9.0
```

In [6]:

```
!pip install rouge
```

```
Collecting rouge
  Downloading rouge-1.0.1-py3-none-any.whl (13 kB)
Requirement already satisfied: six in c:\programdata\anaconda3\lib\site-pack
ages (from rouge) (1.16.0)
Installing collected packages: rouge
Successfully installed rouge-1.0.1
```

In [7]:

```
# We have to remove the first line of each file and store it as an array of headings.
import os
from_dir="bbcspot/athletics"
to_dir="bbcspot/athletics/pre"

for _,_,d in os.walk(from_dir):
    names=d
del(d)

first_line=[]
for i in range(len(names)):
    file1 = from_dir+'/'+names[i]
    file2 = to_dir+'/'+names[i]
    with open(file1, 'r') as fin:
        first = fin.readline()
        first_line.append(first)
        data = fin.read().splitlines(True)
    with open(file2, 'w') as fout:
        fout.writelines(data[1:])

first_line
```

Out[7]:

```
['Claxton hunting first major medal\n',
'O'Sullivan could run in Worlds\n",
'Greene sets sights on world title\n',
'IAAF launches fight against drugs\n',
'Dibaba breaks 5,000m world record\n',
'Isinbayeva claims new world best\n',
'O'Sullivan commits to Dublin race\n",
'Hansen 'delays return until 2006'\n",
'Off-colour Gardener storms to win\n',
'Collins to compete in Birmingham\n',
'Radcliffe yet to answer GB call\n',
'Edwards tips Idowu for Euro gold\n',
"Kenya lift Chepkemei's suspension\n",
'McIlroy aiming for Madrid title\n',
'UK Athletics agrees new kit deal\n',
'Verdict delay for Greek sprinters\n',
'Call for Kenteris to be cleared\n',
'Merritt close to indoor 400m mark\n',
'London hope over Chepkemei\n',
'Edwards tips Idowu for Euro gold\n',
'Chepkemei hit by big ban\n',
'Holmes secures comeback victory\n',
'Greek pair attend drugs hearing\n',
'Chepkemei joins Edinburgh line-up\n',
'El Guerrouj targets cross country\n',
'Collins appeals against drugs ban\n',
'Holmes feted with further honour\n',
'Campbell to extend sprint career\n',
'Isinbayeva heads for Birmingham\n',
'2004: An Irish Athletics Year\n',
'Sprinter Walker quits athletics\n',
'Bekele sets sights on world mark\n',
'Greek sprinters suspended by IAAF\n',
'Athens memories soar above lows\n',
'Collins banned in landmark case\n',
```

'Jones doping probe begins\n',  
'Johnson accuses British sprinters\n',  
'Radcliffe proves doubters wrong\n',  
"Jones medals 'must go if guilty'\n",  
'Radcliffe tackles marathon tasks\n',  
'Gatlin and Hayes win Owen awards\n',  
'Radcliffe enjoys winning comeback\n',  
'Collins named UK Athletics chief\n',  
'White admits to Balco drugs link\n',  
'Trial date is set for Balco case\n',  
'Radcliffe eyes hard line on drugs\n',  
'Kenteris denies faking road crash\n',  
'Greek pair set for hearing\n',  
'Jones files Conte lawsuit\n',  
"IAAF awaits Greek pair's response\n",  
'IAAF to rule on Greek sprint pair\n',  
'Holmes starts 2005 with GB events\n',  
'Gebrselassie in London triple bid\n',  
'Thanou bullish over drugs hearing\n',  
'Holmes is hit by hamstring injury\n',  
'Kluft impressed by Sotherton form\n',  
'What now for Kelly Holmes?\n',  
'Gardener battles to narrow win\n',  
'European medal chances improve\n',  
'McIlroy wins 800m indoor title\n',  
'Johnson edges out rival Sotherton\n',  
"Brizzel to run AAA's in Sheffield\n",  
'Holmes urged to compete at Worlds\n',  
'McIlroy continues winning streak\n',  
'Radcliffe will compete in London\n',  
'Britain boosted by Holmes double\n',  
'Lewis-Francis eyeing world gold\n',  
'GB select Holmes for double bid\n',  
'Holmes back on form in Birmingham\n',  
'Record fails to lift lacklustre meet\n',  
'Tomlinson stays focused on Europe\n',  
'Running around the Olympics\n',  
'Johnson uncertain about Euro bid\n',  
'Pavey focuses on indoor success\n',  
'Gardener wins double in Glasgow\n',  
'African double in Edinburgh\n',  
'Yelling takes Cardiff hat-trick\n',  
'Butler strikes gold in Spain\n',  
'Wada will appeal against ruling\n',  
"Greek sprinters 'won't run again'\n",  
'IAAF will contest Greek decision\n',  
'Greek duo cleared in doping case\n',  
'Johnson too strong for GB runners\n',  
'Holmes starts 2005 with GB events\n',  
'Tulu to appear at Caledonian run\n',  
'Disappointed Scott in solid start\n',  
'GB quartet get cross country call\n',  
'Jones files lawsuit against Conte\n',  
'Thanou desperate to make return\n',  
'Relay squad thrilled with honours\n',  
'Balco case trial date pushed back\n',  
'Collins calls for Chambers return\n',  
'Costin aims for comeback in 2006\n',  
'Lewis-Francis turns to Christie\n',  
'Greene to compete in Birmingham\n',  
'Kluft playing down record chance\n',

```
'Pittman drops coach for UK base\n',  
'Lewis-Francis shakes off injury\n',  
'Holmes facing fine over trials\n',  
'Freeman considers return to track\n',  
'Event devised to nurture athletes\n']
```

In [8]:

```
# remove \n from the end of each heading
headings=[]
for x in first_line:
    y=x.strip()
    headings.append(y)
headings
```

Out[8]:

```
['Claxton hunting first major medal',
'O'Sullivan could run in Worlds",
'Greene sets sights on world title',
'IAAF launches fight against drugs',
'Dibaba breaks 5,000m world record',
'Isinbayeva claims new world best',
'O'Sullivan commits to Dublin race",
'Hansen 'delays return until 2006"',
'Off-colour Gardener storms to win',
'Collins to compete in Birmingham',
'Radcliffe yet to answer GB call',
'Edwards tips Idowu for Euro gold',
'Kenya lift Chepkemei's suspension",
'McIlroy aiming for Madrid title',
'UK Athletics agrees new kit deal',
'Verdict delay for Greek sprinters',
'Call for Kenteris to be cleared',
'Merritt close to indoor 400m mark',
'London hope over Chepkemei',
'Edwards tips Idowu for Euro gold',
'Chepkemei hit by big ban',
'Holmes secures comeback victory',
'Greek pair attend drugs hearing',
'Chepkemei joins Edinburgh line-up',
'El Guerrouj targets cross country',
'Collins appeals against drugs ban',
'Holmes feted with further honour',
'Campbell to extend sprint career',
'Isinbayeva heads for Birmingham',
'2004: An Irish Athletics Year',
'Sprinter Walker quits athletics',
'Bekele sets sights on world mark',
'Greek sprinters suspended by IAAF',
'Athens memories soar above lows',
'Collins banned in landmark case',
'Jones doping probe begins',
'Johnson accuses British sprinters',
'Radcliffe proves doubters wrong',
"Jones medals 'must go if guilty'",
'Radcliffe tackles marathon tasks',
'Gatlin and Hayes win Owen awards',
'Radcliffe enjoys winning comeback',
'Collins named UK Athletics chief',
'White admits to Balco drugs link',
'Trial date is set for Balco case',
'Radcliffe eyes hard line on drugs',
'Kenteris denies faking road crash',
'Greek pair set for hearing',
'Jones files Conte lawsuit',
"IAAF awaits Greek pair's response",
```

'IAAF to rule on Greek sprint pair',  
'Holmes starts 2005 with GB events',  
'Gebrselassie in London triple bid',  
'Thanou bullish over drugs hearing',  
'Holmes is hit by hamstring injury',  
'Kluft impressed by Sotherton form',  
'What now for Kelly Holmes?',  
'Gardener battles to narrow win',  
'European medal chances improve',  
'McIlroy wins 800m indoor title',  
'Johnson edges out rival Sotherton',  
"Brizzel to run AAA's in Sheffield",  
'Holmes urged to compete at Worlds',  
'McIlroy continues winning streak',  
'Radcliffe will compete in London',  
'Britain boosted by Holmes double',  
'Lewis-Francis eyeing world gold',  
'GB select Holmes for double bid',  
'Holmes back on form in Birmingham',  
'Record fails to lift lacklustre meet',  
'Tomlinson stays focused on Europe',  
'Running around the Olympics',  
'Johnson uncertain about Euro bid',  
'Pavey focuses on indoor success',  
'Gardener wins double in Glasgow',  
'African double in Edinburgh',  
'Yelling takes Cardiff hat-trick',  
'Butler strikes gold in Spain',  
'Wada will appeal against ruling',  
"Greek sprinters 'won't run again'",  
'IAAF will contest Greek decision',  
'Greek duo cleared in doping case',  
'Johnson too strong for GB runners',  
'Holmes starts 2005 with GB events',  
'Tulu to appear at Caledonian run',  
'Disappointed Scott in solid start',  
'GB quartet get cross country call',  
'Jones files lawsuit against Conte',  
'Thanou desperate to make return',  
'Relay squad thrilled with honours',  
'Balco case trial date pushed back',  
'Collins calls for Chambers return',  
'Costin aims for comeback in 2006',  
'Lewis-Francis turns to Christie',  
'Greene to compete in Birmingham',  
'Kluft playing down record chance',  
'Pittman drops coach for UK base',  
'Lewis-Francis shakes off injury',  
'Holmes facing fine over trials',  
'Freeman considers return to track',  
'Event devised to nurture athletes'

In [ ]:

In [ ]:



# Luhn : words with higher frequency are more important

In [9]:

```
from sumy.parsers.plaintext import PlaintextParser
from sumy.nlp.tokenizers import Tokenizer
from sumy.summarizers.luhn import LuhnSummarizer

predicted_summary=[]

from_dir="bbcspot/athletics/pre"
for _,_,d in os.walk(from_dir):
    names=d
del(d)

summarizer = LuhnSummarizer()
for i in range(len(names)):
    data = from_dir+'/'+names[i]
    parser = PlaintextParser.from_file(data, Tokenizer("english")) # OR from_string
    #Summarize the document with 1 sentence
    summary = summarizer(parser.document, 2)
    predicted_summary.append(summary)

luhn_summary=[]
for x in predicted_summary:
    #print(x)
    l = ' '.join([str(s) for s in list(x)])
    #print("-----")
    luhn_summary.append(l)

from rouge import Rouge
rouge = Rouge()
luhn_scores = rouge.get_scores(headings, luhn_summary, avg=True)
luhn_scores
```

Out[9]:

```
{'rouge-1': {'r': 0.04042365546471084,
  'p': 0.31834040546911835,
  'f': 0.07129062753628958},
 'rouge-2': {'r': 0.0024831712360428473,
  'p': 0.03003300330033003,
  'f': 0.004569432678846581},
 'rouge-l': {'r': 0.036389738611511416,
  'p': 0.28597359735973593,
  'f': 0.06415620965508068}}
```

In [ ]:

In [ ]:

In [ ]:

# LSA : apply SVD on TF-IDF

In [10]:

```
from sumy.parsers.plaintext import PlaintextParser
from sumy.nlp.tokenizers import Tokenizer
from sumy.summarizers.lsa import LsaSummarizer

predicted_summary=[]

from_dir="bbcspot/athletics/pre"
for _,_,d in os.walk(from_dir):
    names=d
del(d)

summarizer = LsaSummarizer()
for i in range(len(names)):
    data = from_dir+'/' +names[i]
    parser = PlaintextParser.from_file(data, Tokenizer("english")) # OR from_string
    #Summarize the document with 1 sentence
    summary = summarizer(parser.document, 2)
    predicted_summary.append(summary)

lsa_summary=[]
for x in predicted_summary:
    #print(x)
    l = ' '.join([str(s) for s in list(x)])
    #print("-----")
    lsa_summary.append(l)

from rouge import Rouge
rouge = Rouge()
lsa_scores = rouge.get_scores(headings, lsa_summary, avg=True)
lsa_scores
```

Out[10]:

```
{'rouge-1': {'r': 0.04798039668269262,
  'p': 0.3393210749646391,
  'f': 0.08364013802310771},
 'rouge-2': {'r': 0.004530172074927196,
  'p': 0.03663366336633664,
  'f': 0.008041751503745689},
 'rouge-l': {'r': 0.04483938863359544,
  'p': 0.31546440358321526,
  'f': 0.07811126051920694}}
```

In [ ]:

In [ ]:

In [ ]:

# LexRank : cosine-similarity b/w TF-IDF vectors of sentences

In [11]:

```
from sumy.parsers.plaintext import PlaintextParser
from sumy.nlp.tokenizers import Tokenizer
from sumy.summarizers.lex_rank import LexRankSummarizer

predicted_summary=[]

from_dir="bbcspot/athletics/pre"
for _,_,d in os.walk(from_dir):
    names=d
del(d)

summarizer = LexRankSummarizer()
for i in range(len(names)):
    data = from_dir+'/'+names[i]
    parser = PlaintextParser.from_file(data, Tokenizer("english")) # OR from_string
    #Summarize the document with 1 sentence
    summary = summarizer(parser.document, 2)
    predicted_summary.append(summary)

lexrank_summary=[]
for x in predicted_summary:
    #print(x)
    l = ' '.join([str(s) for s in list(x)])
    #print("-----")
    lexrank_summary.append(l)

from rouge import Rouge
rouge = Rouge()
lexrank_scores = rouge.get_scores(headings, lexrank_summary, avg=True)
lexrank_scores

#Lexrank_summary
```

Out[11]:

```
{'rouge-1': {'r': 0.04471282079392965,
  'p': 0.31907119283356894,
  'f': 0.07786822233809264},
 'rouge-2': {'r': 0.0040674750094273765,
  'p': 0.04042904290429043,
  'f': 0.00733061507001175},
 'rouge-l': {'r': 0.04181981847825936,
  'p': 0.2992456388495991,
  'f': 0.07286031489581717}}
```

In [ ]:

In [ ]:

# TextRank : common words in sentences / sum of sentence lengths

In [12]:

```
from sumy.parsers.plaintext import PlaintextParser
from sumy.nlp.tokenizers import Tokenizer
from sumy.summarizers.text_rank import TextRankSummarizer

predicted_summary=[]

from_dir="bbcspot/athletics/pre"
for _,_,d in os.walk(from_dir):
    names=d
del(d)

summarizer = TextRankSummarizer()
for i in range(len(names)):
    data = from_dir+'/'+names[i]
    parser = PlaintextParser.from_file(data, Tokenizer("english")) # OR from_string
    #Summarize the document with 1 sentence
    summary = summarizer(parser.document, 2)
    predicted_summary.append(summary)

textrank_summary=[]
for x in predicted_summary:
    #print(x)
    l = ' '.join([str(s) for s in list(x)])
    #print("-----")
    textrank_summary.append(l)

from rouge import Rouge
rouge = Rouge()
textrank_scores = rouge.get_scores(headings, textrank_summary, avg=True)
textrank_scores
```

Out[12]:

```
{'rouge-1': {'r': 0.033486286317375456,
  'p': 0.2805044790193304,
  'f': 0.05943549457446562},
 'rouge-2': {'r': 0.001844709160098319,
  'p': 0.025082508250825083,
  'f': 0.0034272032965901294},
 'rouge-l': {'r': 0.030743250171635234,
  'p': 0.2596181046676095,
  'f': 0.05463207254328962}}
```

In [ ]:

In [ ]:

In [ ]:

## SumBasic : select sentences with highest probability words

In [13]:

```

from sumy.parsers.plaintext import PlaintextParser
from sumy.nlp.tokenizers import Tokenizer
from sumy.summarizers.sum_basic import SumBasicSummarizer

predicted_summary=[]

from_dir="bbcsport/athletics/pre"
for _,_,d in os.walk(from_dir):
    names=d
    del(d)

summarizer = SumBasicSummarizer()
for i in range(len(names)):
    data = from_dir+'/' +names[i]
    parser = PlaintextParser.from_file(data, Tokenizer("english")) # OR from_string
    #Summarize the document with 1 sentence
    summary = summarizer(parser.document, 2)
    predicted_summary.append(summary)

sumbasic_summary=[]
for x in predicted_summary:
    #print(x)
    l = ' '.join([str(s) for s in list(x)])
    #print("-----")
    sumbasic_summary.append(l)

from rouge import Rouge
rouge = Rouge()
sumbasic_scores = rouge.get_scores(headings, sumbasic_summary, avg=True)
sumbasic_scores

```

Out[13]:

```

{'rouge-1': {'r': 0.039457139095775244,
  'p': 0.23661008958038646,
  'f': 0.06668290873211041},
 'rouge-2': {'r': 0.002062908993602063,
  'p': 0.01881188118811881,
  'f': 0.0037120868651838545},
 'rouge-l': {'r': 0.03767687610095691,
  'p': 0.22652050919377636,
  'f': 0.0636826025713449}}

```

In [ ]:

# KL-Sum

In [14]:

```
from sumy.parsers.plaintext import PlaintextParser
from sumy.nlp.tokenizers import Tokenizer
from sumy.summarizers.kl import KLSummarizer

predicted_summary=[]

from_dir="bbcsport/athletics/pre"
for _,_,d in os.walk(from_dir):
    names=d
del(d)

summarizer = KLSummarizer()
for i in range(len(names)):
    data = from_dir+'/' +names[i]
    parser = PlaintextParser.from_file(data, Tokenizer("english")) # OR from_string
    #Summarize the document with 1 sentence
    summary = summarizer(parser.document, 2)
    predicted_summary.append(summary)

klsum_summary=[]
for x in predicted_summary:
    #print(x)
    l = ' '.join([str(s) for s in list(x)])
    #print("-----")
    klsum_summary.append(l)

from rouge import Rouge
rouge = Rouge()
klsum_scores = rouge.get_scores(headings, klsum_summary, avg=True)
klsum_scores
```

Out[14]:

```
{'rouge-1': {'r': 0.033262182145753064,
 'p': 0.22621404997642616,
 'f': 0.05759996996935777},
 'rouge-2': {'r': 0.001535897368017345,
 'p': 0.015676567656765675,
 'f': 0.0027836404575433808},
 'rouge-l': {'r': 0.03178247373535158,
 'p': 0.21810466760961808,
 'f': 0.055105936362539}}
```

In [ ]:

In [ ]:

In [ ]:

# Reduction

In [15]:

```
from sumy.parsers.plaintext import PlaintextParser
from sumy.nlp.tokenizers import Tokenizer
from sumy.summarizers.reduction import ReductionSummarizer

predicted_summary=[]

from_dir="bbcsport/athletics/pre"
for _,_,d in os.walk(from_dir):
    names=d
del(d)

summarizer = ReductionSummarizer()
for i in range(len(names)):
    data = from_dir+'/' +names[i]
    parser = PlaintextParser.from_file(data, Tokenizer("english")) # OR from_string
    #Summarize the document with 1 sentence
    summary = summarizer(parser.document, 2)
    predicted_summary.append(summary)

reduction_summary=[]
for x in predicted_summary:
    #print(x)
    l = ' '.join([str(s) for s in list(x)])
    #print("-----")
    reduction_summary.append(l)

from rouge import Rouge
rouge = Rouge()
reduction_scores = rouge.get_scores(headings, reduction_summary, avg=True)
reduction_scores
```

Out[15]:

```
{'rouge-1': {'r': 0.03507088918513493,
 'p': 0.28017444601603003,
 'f': 0.061941454739753696},
 'rouge-2': {'r': 0.0018683204242454015,
 'p': 0.025082508250825083,
 'f': 0.003465332910646471},
 'rouge-l': {'r': 0.03270178420714569,
 'p': 0.2645685997171145,
 'f': 0.05785466077435912}}
```

# Done

In [ ]:

In [ ]:

In [16]:

```
str1 = "\n Starbucks has the best coffee \n"  
newstr = str1.strip()  
print(newstr)
```

Starbucks has the best coffee



In [17]:

```
from path import Path

headings=[]
documents_dir = Path('bbcsport/athletics')

for file_path in documents_dir.files('*.txt'):
    with file_path.open(mode='rt', encoding='utf-8') as fp:
        headings.append(fp.readline())

summ=[]
for x in headings:
    summ.append(x.replace("\n", ""))
summ
```

Out[17]:

```
['Claxton hunting first major medal',
'O'Sullivan could run in Worlds',
'Greene sets sights on world title',
'IAAF launches fight against drugs',
'Dibaba breaks 5,000m world record',
'Isinbayeva claims new world best',
'O'Sullivan commits to Dublin race',
'Hansen 'delays return until 2006'',
'Off-colour Gardener storms to win',
'Collins to compete in Birmingham',
'Radcliffe yet to answer GB call',
'Edwards tips Idowu for Euro gold',
'Kenya lift Chepkemei's suspension',
'McIlroy aiming for Madrid title',
'UK Athletics agrees new kit deal',
'Verdict delay for Greek sprinters',
'Call for Kenteris to be cleared',
'Merritt close to indoor 400m mark',
'London hope over Chepkemei',
'Edwards tips Idowu for Euro gold',
'Chepkemei hit by big ban',
'Holmes secures comeback victory',
'Greek pair attend drugs hearing',
'Chepkemei joins Edinburgh line-up',
'El Guerrouj targets cross country',
'Collins appeals against drugs ban',
'Holmes feted with further honour',
'Campbell to extend sprint career',
'Isinbayeva heads for Birmingham',
'2004: An Irish Athletics Year',
'Sprinter Walker quits athletics',
'Bekele sets sights on world mark',
'Greek sprinters suspended by IAAF',
'Athens memories soar above lows',
'Collins banned in landmark case',
'Jones doping probe begins',
'Johnson accuses British sprinters',
'Radcliffe proves doubters wrong',
'Jones medals 'must go if guilty'',
'Radcliffe tackles marathon tasks',
'Gatlin and Hayes win Owen awards',
'Radcliffe enjoys winning comeback',
'Collins named UK Athletics chief',
```

'White admits to Balco drugs link',  
'Trial date is set for Balco case',  
'Radcliffe eyes hard line on drugs',  
'Kenteris denies faking road crash',  
'Greek pair set for hearing',  
'Jones files Conte lawsuit',  
"IAAF awaits Greek pair's response",  
'IAAF to rule on Greek sprint pair',  
'Holmes starts 2005 with GB events',  
'Gebrselassie in London triple bid',  
'Thanou bullish over drugs hearing',  
'Holmes is hit by hamstring injury',  
'Kluft impressed by Sotherton form',  
'What now for Kelly Holmes?',  
'Gardener battles to narrow win',  
'European medal chances improve',  
'McIlroy wins 800m indoor title',  
'Johnson edges out rival Sotherton',  
"Brizzel to run AAA's in Sheffield",  
'Holmes urged to compete at Worlds',  
'McIlroy continues winning streak',  
'Radcliffe will compete in London',  
'Britain boosted by Holmes double',  
'Lewis-Francis eyeing world gold',  
'GB select Holmes for double bid',  
'Holmes back on form in Birmingham',  
'Record fails to lift lacklustre meet',  
'Tomlinson stays focused on Europe',  
'Running around the Olympics',  
'Johnson uncertain about Euro bid',  
'Pavey focuses on indoor success',  
'Gardener wins double in Glasgow',  
'African double in Edinburgh',  
'Yelling takes Cardiff hat-trick',  
'Butler strikes gold in Spain',  
'Wada will appeal against ruling',  
"Greek sprinters 'won't run again'",  
'IAAF will contest Greek decision',  
'Greek duo cleared in doping case',  
'Johnson too strong for GB runners',  
'Holmes starts 2005 with GB events',  
'Tulu to appear at Caledonian run',  
'Disappointed Scott in solid start',  
'GB quartet get cross country call',  
'Jones files lawsuit against Conte',  
'Thanou desperate to make return',  
'Relay squad thrilled with honours',  
'Balco case trial date pushed back',  
'Collins calls for Chambers return',  
'Costin aims for comeback in 2006',  
'Lewis-Francis turns to Christie',  
'Greene to compete in Birmingham',  
'Kluft playing down record chance',  
'Pittman drops coach for UK base',  
'Lewis-Francis shakes off injury',  
'Holmes facing fine over trials',  
'Freeman considers return to track',  
'Event devised to nurture athletes']

In [18]:

```
from path import Path

documents = []
headings=[]
documents_dir = Path('bbcsport/athletics')

for file_path in documents_dir.files('*.txt'):
    with file_path.open(mode='rt', encoding='utf-8') as fp:
        headings.append(fp.readline())

for file_path in documents_dir.files('*.txt'):
    with file_path.open(mode='rt', encoding='utf-8') as fp:
        documents.append(fp.readlines()[1:])
```

In [19]:

headings

Out[19]:

```
['Claxton hunting first major medal\n',  
"O'Sullivan could run in Worlds\n",  
'Greene sets sights on world title\n',  
'IAAF launches fight against drugs\n',  
'Dibaba breaks 5,000m world record\n',  
'Isinbayeva claims new world best\n',  
"O'Sullivan commits to Dublin race\n",  
"Hansen 'delays return until 2006'\n",  
'Off-colour Gardener storms to win\n',  
'Collins to compete in Birmingham\n',  
'Radcliffe yet to answer GB call\n',  
'Edwards tips Idowu for Euro gold\n',  
"Kenya lift Chepkemei's suspension\n",  
'McIlroy aiming for Madrid title\n',  
'UK Athletics agrees new kit deal\n',  
'Verdict delay for Greek sprinters\n',  
'Call for Kenteris to be cleared\n',  
'Merritt close to indoor 400m mark\n',  
'London hope over Chepkemei\n',  
'Edwards tips Idowu for Euro gold\n',  
'Chepkemei hit by big ban\n',  
'Holmes secures comeback victory\n',  
'Greek pair attend drugs hearing\n',  
'Chepkemei joins Edinburgh line-up\n',  
'El Guerrouj targets cross country\n',  
'Collins appeals against drugs ban\n',  
'Holmes feted with further honour\n',  
'Campbell to extend sprint career\n',  
'Isinbayeva heads for Birmingham\n',  
'2004: An Irish Athletics Year\n',  
'Sprinter Walker quits athletics\n',  
'Bekele sets sights on world mark\n',  
'Greek sprinters suspended by IAAF\n',  
'Athens memories soar above lows\n',  
'Collins banned in landmark case\n',  
'Jones doping probe begins\n',  
'Johnson accuses British sprinters\n',  
'Radcliffe proves doubters wrong\n',  
"Jones medals 'must go if guilty'\n",  
'Radcliffe tackles marathon tasks\n',  
'Gatlin and Hayes win Owen awards\n',  
'Radcliffe enjoys winning comeback\n',  
'Collins named UK Athletics chief\n',  
'White admits to Balco drugs link\n',  
'Trial date is set for Balco case\n',  
'Radcliffe eyes hard line on drugs\n',  
'Kenteris denies faking road crash\n',  
'Greek pair set for hearing\n',  
'Jones files Conte lawsuit\n',  
"IAAF awaits Greek pair's response\n",  
'IAAF to rule on Greek sprint pair\n',  
'Holmes starts 2005 with GB events\n',  
'Gebrselassie in London triple bid\n',  
'Thanou bullish over drugs hearing\n',  
'Holmes is hit by hamstring injury\n',
```

'Kluft impressed by Sotherton form\n',  
'What now for Kelly Holmes?\n',  
'Gardener battles to narrow win\n',  
'European medal chances improve\n',  
'McIlroy wins 800m indoor title\n',  
'Johnson edges out rival Sotherton\n',  
"Brizzel to run AAA's in Sheffield\n",  
'Holmes urged to compete at Worlds\n',  
'McIlroy continues winning streak\n',  
'Radcliffe will compete in London\n',  
'Britain boosted by Holmes double\n',  
'Lewis-Francis eyeing world gold\n',  
'GB select Holmes for double bid\n',  
'Holmes back on form in Birmingham\n',  
'Record fails to lift lacklustre meet\n',  
'Tomlinson stays focused on Europe\n',  
'Running around the Olympics\n',  
'Johnson uncertain about Euro bid\n',  
'Pavey focuses on indoor success\n',  
'Gardener wins double in Glasgow\n',  
'African double in Edinburgh\n',  
'Yelling takes Cardiff hat-trick\n',  
'Butler strikes gold in Spain\n',  
'Wada will appeal against ruling\n',  
"Greek sprinters 'won't run again'\n",  
'IAAF will contest Greek decision\n',  
'Greek duo cleared in doping case\n',  
'Johnson too strong for GB runners\n',  
'Holmes starts 2005 with GB events\n',  
'Tulu to appear at Caledonian run\n',  
'Disappointed Scott in solid start\n',  
'GB quartet get cross country call\n',  
'Jones files lawsuit against Conte\n',  
'Thanou desperate to make return\n',  
'Relay squad thrilled with honours\n',  
'Balco case trial date pushed back\n',  
'Collins calls for Chambers return\n',  
'Costin aims for comeback in 2006\n',  
'Lewis-Francis turns to Christie\n',  
'Greene to compete in Birmingham\n',  
'Kluft playing down record chance\n',  
'Pittman drops coach for UK base\n',  
'Lewis-Francis shakes off injury\n',  
'Holmes facing fine over trials\n',  
'Freeman considers return to track\n',  
'Event devised to nurture athletes\n']

In [20]:

```
summ=[]  
for x in headings:  
    summ.append(x.replace("\n", ""))  
summ
```

Out[20]:

```
['Claxton hunting first major medal',  
"O'Sullivan could run in Worlds",  
'Greene sets sights on world title',  
'IAAF launches fight against drugs',  
'Dibaba breaks 5,000m world record',  
'Isinbayeva claims new world best',  
"O'Sullivan commits to Dublin race",  
"Hansen 'delays return until 2006'",  
'Off-colour Gardener storms to win',  
'Collins to compete in Birmingham',  
'Radcliffe yet to answer GB call',  
'Edwards tips Idowu for Euro gold',  
"Kenya lift Chepkemei's suspension",  
'McIlroy aiming for Madrid title',  
'UK Athletics agrees new kit deal',  
'Verdict delay for Greek sprinters',  
'Call for Kenteris to be cleared',  
'Merritt close to indoor 400m mark',  
'London hope over Chepkemei',  
'Edwards tips Idowu for Euro gold',  
'Chepkemei hit by big ban',  
'Holmes secures comeback victory',  
'Greek pair attend drugs hearing',  
'Chepkemei joins Edinburgh line-up',  
'El Guerrouj targets cross country',  
'Collins appeals against drugs ban',  
'Holmes feted with further honour',  
'Campbell to extend sprint career',  
'Isinbayeva heads for Birmingham',  
'2004: An Irish Athletics Year',  
'Sprinter Walker quits athletics',  
'Bekele sets sights on world mark',  
'Greek sprinters suspended by IAAF',  
'Athens memories soar above lows',  
'Collins banned in landmark case',  
'Jones doping probe begins',  
'Johnson accuses British sprinters',  
'Radcliffe proves doubters wrong',  
"Jones medals 'must go if guilty'",  
'Radcliffe tackles marathon tasks',  
'Gatlin and Hayes win Owen awards',  
'Radcliffe enjoys winning comeback',  
'Collins named UK Athletics chief',  
'White admits to Balco drugs link',  
'Trial date is set for Balco case',  
'Radcliffe eyes hard line on drugs',  
'Kenteris denies faking road crash',  
'Greek pair set for hearing',  
'Jones files Conte lawsuit',  
"IAAF awaits Greek pair's response",  
'IAAF to rule on Greek sprint pair',  
'Holmes starts 2005 with GB events',
```

'Gebrselassie in London triple bid',  
'Thanou bullish over drugs hearing',  
'Holmes is hit by hamstring injury',  
'Kluft impressed by Sotherton form',  
'What now for Kelly Holmes?',  
'Gardener battles to narrow win',  
'European medal chances improve',  
'McIlroy wins 800m indoor title',  
'Johnson edges out rival Sotherton',  
"Brizzel to run AAA's in Sheffield",  
'Holmes urged to compete at Worlds',  
'McIlroy continues winning streak',  
'Radcliffe will compete in London',  
'Britain boosted by Holmes double',  
'Lewis-Francis eyeing world gold',  
'GB select Holmes for double bid',  
'Holmes back on form in Birmingham',  
'Record fails to lift lacklustre meet',  
'Tomlinson stays focused on Europe',  
'Running around the Olympics',  
'Johnson uncertain about Euro bid',  
'Pavey focuses on indoor success',  
'Gardener wins double in Glasgow',  
'African double in Edinburgh',  
'Yelling takes Cardiff hat-trick',  
'Butler strikes gold in Spain',  
'Wada will appeal against ruling',  
"Greek sprinters 'won't run again'",  
'IAAF will contest Greek decision',  
'Greek duo cleared in doping case',  
'Johnson too strong for GB runners',  
'Holmes starts 2005 with GB events',  
'Tulu to appear at Caledonian run',  
'Disappointed Scott in solid start',  
'GB quartet get cross country call',  
'Jones files lawsuit against Conte',  
'Thanou desperate to make return',  
'Relay squad thrilled with honours',  
'Balco case trial date pushed back',  
'Collins calls for Chambers return',  
'Costin aims for comeback in 2006',  
'Lewis-Francis turns to Christie',  
'Greene to compete in Birmingham',  
'Kluft playing down record chance',  
'Pittman drops coach for UK base',  
'Lewis-Francis shakes off injury',  
'Holmes facing fine over trials',  
'Freeman considers return to track',  
'Event devised to nurture athletes']

In [21]:

documents[0]

Out[21]:

```
['\n',
 "British hurdler Sarah Claxton is confident she can win her first major medal at next month's European Indoor Championships in Madrid.\n",
 '\n',
 'The 25-year-old has already smashed the British record over 60m hurdles twice this season, setting a new mark of 7.96 seconds to win the AAAs title. "I am quite confident," said Claxton. "But I take each race as it comes. "As long as I keep up my training but not do too much I think there is a chance of a medal." Claxton has won the national 60m hurdles title for the past three years but has struggled to translate her domestic success to the international stage. Now, the Scotland-born athlete owns the equal fifth-fastest time in the world this year. And at last week\'s Birmingham Grand Prix, Claxton left European medal favourite Russian Irina Shevchenko trailing in sixth spot.\n',
 '\n',
 'For the first time, Claxton has only been preparing for a campaign over the hurdles - which could explain her leap in form. In previous seasons, the 25-year-old also contested the long jump but since moving from Colchester to London she has re-focused her attentions. Claxton will see if her new training regime pays dividends at the European Indoors which take place on 5-6 March.\n']
```

In [22]:

```
text=[]
for x in documents:
    print(x)
    text.append(x)
#text
```

```
['\n', "The IAAF - athletics' world governing body - has met anti-doping officials, coaches and athletes to co-ordinate the fight against drugs in sport.\n", '\n', 'Two task forces have been set up to examine doping and nutrition issues. It was also agreed that a programme to "de-mystify" the issue to athletes, the public and the media was a priority. "Nothing was decided to change things - it was more to have a forum of the stakeholders allowing them to express themselves," said an IAAF spokesman. "Getting everyone together gave us a lot of food for thought." About 60 people attended Sunday\'s meeting in Monaco, including IAAF chief Lamine Diack and Namibian athlete Frankie Fredericks, now a member of the Athletes\' Commission. "I am very happy to see you all, members of the athletics family, respond positively to the IAAF call to sit together and discuss what more we can do in the fight against doping," said Diack. "We are the leading Federation in this field and it is our duty to keep our sport clean." The two task forces will report back to the IAAF Council, at its April meeting in Qatar.\n']
['\n', "Ethiopia's Tirunesh Dibaba set a new world record in winning the women's 5,000m at the Boston Indoor Games.\n", '\n', 'Dibaba won in 14 minutes 32.93 seconds to erase the previous world indoor mark of 14:39.29 set by another Ethiopian, Berhane Adere, in Stuttgart last year. But compatriot
```



In [23]:

```
docs=[]
for x in text:
    para=[]
    for y in x:
        if(y=="\n"):
            pass
        else:
            para.append(y.replace("\n", ""))
            #y.strip()
            #para.append(y)
    docs.append(para)
docs
```

Out[23]:

["British hurdler Sarah Claxton is confident she can win her first major medal at next month's European Indoor Championships in Madrid.",

'The 25-year-old has already smashed the British record over 60m hurdles twice this season, setting a new mark of 7.96 seconds to win the AAAs title. "I am quite confident," said Claxton. "But I take each race as it comes. "As long as I keep up my training but not do too much I think there is a chance of a medal." Claxton has won the national 60m hurdles title for the past three years but has struggled to translate her domestic success to the international stage. Now, the Scotland-born athlete owns the equal fifth-fastest time in the world this year. And at last week's Birmingham Grand Prix, Claxton left European medal favourite Russian Irina Shevchenko trailing in sixth spot.',

'For the first time, Claxton has only been preparing for a campaign over the hurdles - which could explain her leap in form. In previous seasons, the 25-year-old also contested the long jump but since moving from Colchester to London she has re-focused her attentions. Claxton will see if her new training regime pays dividends at the European Indoors which take place on 5-6 March.'].

In [24]:

```
merged_text=[]  
for x in docs:  
    y=" ".join(x)  
    merged_text.append(y)  
merged_text
```

Out[24]:

['British hurdler Sarah Claxton is confident she can win her first major medal at next month\'s European Indoor Championships in Madrid. The 25-year-old has already smashed the British record over 60m hurdles twice this season, setting a new mark of 7.96 seconds to win the AAAs title. "I am quite confident," said Claxton. "But I take each race as it comes. "As long as I keep up my training but not do too much I think there is a chance of a medal." Claxton has won the national 60m hurdles title for the past three years but has struggled to translate her domestic success to the international stage. Now, the Scotland-born athlete owns the equal fifth-fastest time in the world this year. And at last week\'s Birmingham Grand Prix, Claxton left European medal favourite Russian Irina Shevchenko trailing in sixth spot. For the first time, Claxton has only been preparing for a campaign over the hurdles - which could explain her leap in form. In previous seasons, the 25-year-old also contested the long jump but since moving from Colchester to London she has re-focused her attentions. Claxton will see if her new training regime pays dividends at the European Indoors which take place on 5-6 March.',  
"Sonia O'Sullivan has indicated that she would like to participate in nex

In [25]:

```
import pandas as pd  
df = pd.DataFrame()  
df["text"] = merged_text  
df["summary"] = summ  
df.to_csv("preprocessed_bbc_athletics.csv")
```

In [26]:

```
for i in range(5):  
    print("News Item #",i)  
    print(merged_text[i])  
    print(summ[i])  
    print()
```

#### News Item # 0

British hurdler Sarah Claxton is confident she can win her first major medal at next month's European Indoor Championships in Madrid. The 25-year-old has already smashed the British record over 60m hurdles twice this season, setting a new mark of 7.96 seconds to win the AAAs title. "I am quite confident," said Claxton. "But I take each race as it comes. "As long as I keep up my training but not do too much I think there is a chance of a medal." Claxton has won the national 60m hurdles title for the past three years but has struggled to translate her domestic success to the international stage. Now, the Scotland-born athlete owns the equal fifth-fastest time in the world this year. And at last week's Birmingham Grand Prix, Claxton left European medal favourite Russian Irina Shevchenko trailing in sixth spot. For the first time, Claxton has only been preparing for a campaign over the hurdles - which could explain her leap in form. In previous seasons, the 25-year-old also contested the long jump but since moving from Colchester to London she has refocused her attentions. Claxton will see if her new training regime pays dividends at the European Indoors which take place on 5-6 March. Claxton hunting first major medal

#### News Item # 1

Sonia O'Sullivan has indicated that she would like to participate in next month's World Cross Country Championships in St Etienne. Athletics Ireland have hinted that the 35-year-old Cobh runner may be included in the official line-up for the event in France on 19-20 March. Provincial teams were selected after last Saturday's Nationals in Santry and will be officially announced this week. O'Sullivan is at present preparing for the London marathon on 17 April. The participation of O'Sullivan, currently training at her base in Australia, would boost the Ireland team who won the bronze three years ago. The first three at Santry last Saturday, Jolene Byrne, Maria McCambridge and Fionnuala Britton, are automatic selections and will most likely form part of the long-course team. O'Sullivan will also take part in the Bupa Great Ireland Run on 9 April in Dublin. O'Sullivan could run in Worlds

#### News Item # 2

Maurice Greene aims to wipe out the pain of losing his Olympic 100m title in Athens by winning a fourth World Championship crown this summer. He had to settle for bronze in Greece behind fellow American Justin Gatlin and Francis Obikwelu of Portugal. "It really hurts to look at that medal. It was my mistake. I lost because of the things I did," said Greene, who races in Birmingham on Friday. "It's never going to happen again. My goal - I'm going to win the worlds." Greene crossed the line just 0.02 seconds behind Gatlin, who won in 9.87 seconds in one of the closest and fastest sprints of all time. But Greene believes he lost the race and his title in the semi-finals. "In my semi-final race, I should have won the race but I was conserving energy. That's when Francis Obikwelu came up and I took third because I didn't know he was there. "I believe that's what put me in lane seven in the final and, while I was in lane seven, I couldn't feel anything in the race. "I just felt like I was running all alone. "I believe if I was in the middle of the race I would have been able to react to people that came ahead of me." Greene was also denied Olympic gold in the 4x100m men's relay when he could not catch Britain's Mark Lewis-Francis on the final leg. The Kansas star is set to go head-to-head with Lewis-Francis again at Friday's Norwich Union Grand Prix. Th

e pair contest the 60m, the distance over which Greene currently holds the world record of 6.39 seconds. He then has another indoor meeting in France before resuming training for the outdoor season and the task of recapturing his world title in Helsinki in August. Greene believes Gatlin will again prove the biggest threat to his ambitions in Finland. But he also admits he faces more than one rival for the world crown. "There's always someone else coming. I think when I was coming up I would say there was me and Ato (Boldon) in the young crowd," Greene said. "Now you've got about five or six young guys coming up at the same time."

Greene sets sights on world title

#### News Item # 3

The IAAF - athletics' world governing body - has met anti-doping officials, coaches and athletes to co-ordinate the fight against drugs in sport. Two task forces have been set up to examine doping and nutrition issues. It was also agreed that a programme to "de-mystify" the issue to athletes, the public and the media was a priority. "Nothing was decided to change things - it was more to have a forum of the stakeholders allowing them to express themselves," said an IAAF spokesman. "Getting everyone together gave us a lot of food for thought." About 60 people attended Sunday's meeting in Monaco, including IAAF chief Lamine Diack and Namibian athlete Frankie Fredericks, now a member of the Athletes' Commission. "I am very happy to see you all, members of the athletics family, respond positively to the IAAF call to sit together and discuss what more we can do in the fight against doping," said Diack. "We are the leading Federation in this field and it is our duty to keep our sport clean." The two task forces will report back to the IAAF Council, at its April meeting in Qatar.

IAAF launches fight against drugs

#### News Item # 4

Ethiopia's Tirunesh Dibaba set a new world record in winning the women's 5,000m at the Boston Indoor Games. Dibaba won in 14 minutes 32.93 seconds to erase the previous world indoor mark of 14:39.29 set by another Ethiopian, Berhane Adera, in Stuttgart last year. But compatriot Kenenisa Bekele's record hopes were dashed when he miscounted his laps in the men's 3,000m and staged his sprint finish a lap too soon. Ireland's Alistair Cragg won in 7:39.89 as Bekele battled to second in 7:41.42. "I didn't want to sit back and get out-kicked," said Cragg. "So I kept on the pace. The plan was to go with 500m to go no matter what, but when Bekele made the mistake that was it. The race was mine." Sweden's Carolina Klüft, the Olympic heptathlon champion, and Slovenia's Jolanda Ceplak had winning performances, too. Klüft took the long jump at 6.63m, while Ceplak easily won the women's 800m in 2:01.52.

Dibaba breaks 5,000m world record

In [ ]: