

Segmentation

QUESTIONS from 01a file:

For this task I used Pragmatic segmenter and NLTK's Punkt segmentor

Pragmatic Segmeter output

he princely state of Mandi was founded by Bahu Sen in 1288 AD.
But Ajabr Sen was the one who founded the historical city of Mandi in 1526 AD.
The Chiefs of Mandi State are said to be the descendants of common ancestors of the Chandervanshi line of Rajputs of Sen dynasty of Bengal who claim to be the descendants of Pandavas of Mahabharata times.
The present District of Mandi was formed with the merger of two princely states Mandi State and Suket (Sundernagar) on 15 April 1948, when the State of Himachal Pradesh was established.
Ajabr Sen assassinated Mandi, it led to the beginning of the modern period of Mandi history.
Old Mandi capital was abandoned, and the site of present capital was acquired.
Some historians put the selection of present capital in the year 1527 AD.
The idol of Madho Rai, the presiding deity of Mandi and incarnation of Lord Vishnu is said to be made by Goldsmith Bhima, in 1648.
Maharaja Ranjit Singh occupied Mandi in 1839 but Ranjit Singh died on 27 June 1839.
It was beginning of the disintegration of Sikh powers in the hills and British power emerged as most powerful.
In 1849, Princely states of Mandi, Suket and Chamba were put under the control of Superintendent 'Cis-Satluj States'.
Lala Lejpat Rai visited Mandi during 1906 to organize the revolutionary activities.
The revolt occurred in Mandi State against the repressive, corrupt and arrogant administration of Raja Bhawani Sen and his Wazir Jiwa Nand Paddha.
This revolt was led by Shobha Ram in 1909.
A Prominent revolutionary of Mandi, Hardev Ram joined a revolutionary band of patriots in 1913.
He went to Shanghai, America and Japan where he met Lala Hardayal and Dr. Mathura Singh.
Another revolutionary of Mandi Bhai Hirda Ram started participating in revolutionary activities in 1914.
He joined Ras Behari Bose, Sant Baba Wasakha Singh Dadehar, Mathura Singh, Kartar Singh Sarba, Bhai Parma Nand and Pingle.
In 1915, on the request of Sarla Devi, wife of Bhai Hirda Ram, the Viceroy converted his death sentence into life imprisonment.
In the same year, Nagchala Dacoity in Mandi state took place.
In Mandi's conspiracy case, Mandi revolutionary Sidhu Kharara was sentenced for seven years in 1917.
Mandi Legislative Council was constituted in 1933.
Swami Purna Nand of Mandi was elected as president of "Himalayan Hill states regional Council" (founded in 1945) & Pt.
Padam Dev became its general secretary.
The session of "Himalayan Hill states regional Council" was held at Mandi from 8 March 1946 to 10 March 1946.
It was also attended by INA activists Dhillon, Sehgal and Shahnawaj.
So finally, Mandi came into existence on 15 April 1948.
Suraj Sen ruled over Mandi at time of Dandana Mahal of Mandi.
(26)Before independence
See also: Mandi State
The Chiefs of Mandi State are said to be the descendants of common ancestors of the Chandervanshi line of rajputs of Sen dynasty of Bengal who claim to be the descendants of Pandavas of Mahabharata times.
According to the earlier accounts, Bahu Sen established an independent territory and was recognized as a Rana or local Chief.
The Raja Ajabr Sen, the nineteenth descendant from Bahu Sen, was considered as an able and benevolent ruler.
He founded the present Mandi city around the 'Bhootnath Temple' in the early 16th century (1588 AD-1534 AD) and shifted his capital to this place from Purani Mandi.
The temple of Trilokinath was also constructed during his reign.
He was succeeded by Chhattar Sen who built a fort in Kalar by name Lakargarh.
The successive rulers of Mandi were Shyan Sen, Sigh Sen, Shamsheer Sen, Surma Sen, Ishwari Sen, Jalam Sen, Balbir Sen and Vijay Sen. Many public utility, services and facilities like schools, hospitals and post offices, etc., were started by Raja Vijay Sen during his time.
He also built large number of palaces and houses.
A mule road from Bajjnath in District Kangra to Sultampur in adjoining District of Kullu, via Mandi and the Victoria suspension Bridge on river Beas were constructed during his time in the year 1877.
Raja Vijay Sen was issue less and left no direct heir.
So in 1897, Bhawani Sen was recognized as his successor.
During his period the city was considerably damaged in the earthquake of 1905.
He was succeeded by Joginder Sen in April 1914 who was, however, vested with full powers of ruling chief in February 1925.
During the period of his minority, Mandi remained under the management of a British officer, and a number of departments were overhauled and reorganised.
Mandi being one of the oldest city is enjoying urban status since 1981.
The Mandi palace popularly known as 'Raj Mahal' as it looked in the past.
The area where the 'Sunken

NLTK Punkt output

he princely state of Mandi was founded by Bahu Sen in 1288 AD.
But Ajabr Sen was the one who founded the historical city of Mandi in 1526 AD.
The Chiefs of Mandi State are said to be the descendants of common ancestors of the Chandervanshi line of Rajputs of Sen dynasty of Bengal who claim to be the descendants of Pandavas of Mahabharata times.
The present District of Mandi was formed with the merger of two princely states Mandi State and Suket (Sundernagar) on 15 April 1948, when the State of Himachal Pradesh was established.
Ajabr Sen assassinated Mandi, it led to the beginning of the modern period of Mandi history.
Old Mandi capital was abandoned, and the site of present capital was acquired.
Some historians put the selection of present capital in the year 1527 AD.
The idol of Madho Rai, the presiding deity of Mandi and incarnation of Lord Vishnu is said to be made by Goldsmith Bhima, in 1648.
Maharaja Ranjit Singh occupied Mandi in 1839 but Ranjit Singh died on 27 June 1839.
It was beginning of the disintegration of Sikh powers in the hills and British power emerged as most powerful.
In 1849, Princely states of Mandi, Suket and Chamba were put under the control of Superintendent 'Cis-Satluj States'.
Lala Lejpat Rai visited Mandi during 1906 to organize the revolutionary activities.
The revolt occurred in Mandi State against the repressive, corrupt and arrogant administration of Raja Bhawani Sen and his Wazir Jiwa Nand Paddha.
This revolt was led by Shobha Ram in 1909.
A Prominent revolutionary of Mandi, Hardev Ram joined a revolutionary band of patriots in 1913.
He went to Shanghai, America and Japan where he met Lala Hardayal and Dr. Mathura Singh.
Another revolutionary of Mandi Bhai Hirda Ram started participating in revolutionary activities in 1914.
He joined Ras Behari Bose, Sant Baba Wasakha Singh Dadehar, Mathura Singh, Kartar Singh Sarba, Bhai Parma Nand and Pingle.
In 1915, on the request of Sarla Devi, wife of Bhai Hirda Ram, the Viceroy converted his death sentence into life imprisonment.
In the same year, Nagchala Dacoity in Mandi state took place.
In Mandi's conspiracy case, Mandi revolutionary Sidhu Kharara was sentenced for seven years in 1917.
Mandi Legislative Council was constituted in 1933.
Swami Purna Nand of Mandi was elected as president of "Himalayan Hill states regional Council" (founded in 1945) & Pt.
Padam Dev became its general secretary.
The session of "Himalayan Hill states regional Council" was held at Mandi from 8 March 1946 to 10 March 1946.
It was also attended by INA activists Dhillon, Sehgal and Shahnawaj.
So finally, Mandi came into existence on 15 April 1948.
Suraj Sen ruled over Mandi at time of Dandana Mahal of Mandi.
(26)Before independenceSee also: Mandi StateThe Chiefs of Mandi State are said to be the descendants of common ancestors of the Chandervanshi line of rajputs of Sen dynasty of Bengal who claim to be the descendants of Pandavas of Mahabharata times.
According to the earlier accounts, Bahu Sen established an independent territory and was recognized as a Rana or local Chief.
The Raja Ajabr Sen, the nineteenth descendant from Bahu Sen, was considered as an able and benevolent ruler.
He founded the present Mandi city around the 'Bhootnath Temple' in the early 16th century (1588 AD-1534 AD) and shifted his capital to this place from Purani Mandi.
The temple of Trilokinath was also constructed during his reign.
He was succeeded by Chhattar Sen who built a fort in Kalar by name Lakargarh.
The successive rulers of Mandi were Shyan Sen, Gaur Sen, Sigh Sen, Shamsheer Sen, Surma Sen, Ishwari Sen, Jalam Sen, Balbir Sen and Vijay Sen. Many public utility, services and facilities like schools, hospitals and post offices, etc., were started by Raja Vijay Sen during his time.
He also built large number of palaces and houses.
A mule road from Bajjnath in District Kangra to Sultampur in adjoining District of Kullu, via Mandi and the Victoria suspension Bridge on river Beas were constructed during his time in the year 1877.
Raja Vijay Sen was issue less and left no direct heir.
So in 1897, Bhawani Sen was recognized as his successor.
During his period the city was considerably damaged in the earthquake of 1905.
He was succeeded by Joginder Sen in April 1914 who was, however, vested with full powers of ruling chief in February 1925.
During the period of his minority, Mandi remained under the management of a British officer, and a number of departments were overhauled and reorganised.
Mandi being one of the oldest city is enjoying urban status since 1981.
The Mandi palace popularly known as 'Raj Mahal' as it looked in the past.
The area where the 'Sunken
(The princely state of Mandi was founded by Bahu Sen in 1288 AD.', 'But Ajabr Sen was the one who founded the historical city of Mandi in 1526 AD.', 'The Chiefs of Mandi State are said to be the descendants of common ancestors of the Chandervanshi line of Rajputs of Sen dynasty of Bengal who claim to be the descendants of Pandavas of Mahabharata times.', 'The present District of Mandi was formed with the merger of two princely states Mandi State and Suket (Sundernagar) on 15 April 1948, when the State of Himachal Pradesh was established.', 'Ajabr Sen assassinated Mandi, it led to the beginning of the modern period of Mandi history.', 'Old Mandi capital was abandoned, and the site of present capital was acquired.', 'Some historians put the selection of present capital in the year 1527 AD.', 'The idol of Madho Rai, the presiding deity of Mandi and incarnation of Lord Vishnu is said to be made by Goldsmith Bhima, in 1648.', 'Maharaja Ranjit Singh occupied Mandi in 1839 but Ranjit Singh died on 27 June 1839.', 'It was beginning of the disintegration of Sikh powers in the hills and British power emerged as most powerful.', 'In 1849, Princely states of Mandi, Suket and Chamba were put under the control of Superintendent 'Cis-Satluj States'.', 'Lala Lejpat Rai visited Mandi during 1906 to organize the revolutionary activities.', 'The revolt occurred in Mandi State against the repressive, corrupt and arrogant administration of Raja Bhawani Sen and his Wazir Jiwa Nand Paddha.', 'This revolt was led by Shobha Ram in 1909.', 'A Prominent revolutionary of Mandi, Hardev Ram joined a revolution ary band of patriots in 1913.', 'He went to Shanghai, America and Japan where he met Lala Hardayal and Dr. Mathura Singh.', 'Another revolutionary of Mandi Bhai Hirda Ram started participating in revolutionary activities in 1914.', 'He joined Ras Behari Bose, Sant Baba Wasakha Singh Dadehar, Mathura Singh, Kartar Singh Sarba, Bhai Parma Nand and Pingle.', 'In 1915, on the request of Sarla Devi, wife of Bhai Hirda Ram, the Viceroy converted his death sent ence into life imprisonment.', 'In the same year, Nagchala Dacoity in Mandi state took place.', 'In Mandi's conspiracy case, Mandi revolutionary Sidhu Kharara was sentenced for seven years in 1917.', 'Mandi Legislative Council wa s constituted in 1933.', 'Swami Purna Nand of Mandi was elected as president of "Himalayan Hill states regional Council" (founded in 1945) & Pt.', 'Padam Dev became its general secretary.', 'The session of "Himalayan Hill states regional Council" was held at Mandi from 8 March 1946 to 10 March 1946.', 'It was also attended by INA activists Dhillon, Sehgal and Shahnawaj.', 'So finally, Mandi came into existence on 15 April 1948.', 'Suraj Sen ruled over Ma ndi at time of Dandana Mahal of Mandi.', '(26)Before independenceSee also: Mandi StateThe Chiefs of Mandi State are said to be the descendants of common ancestors of the Chandervanshi line of rajputs of Sen dynasty of Bengal who claim to be the descendants of Pandavas of Mahabharata times.', 'According to the earlier accounts, Bahu Sen established an independent territory and was recognized as a Rana or local Chief.', 'The Raja Ajabr Sen, the nineteenth descendant from Bahu Sen, was considered as an able and benevolent ruler.', 'He founded the present Mandi city around the 'Bhootnath Temple' in the early 16th century (1588 AD-1534 AD) and shifted his capital to this place from P urani Mandi.', 'The temple of Trilokinath was also constructed during his reign.', 'He was succeeded by Chhattar Sen who built a fort in Kalar by name Lakargarh.', 'The successive rulers of Mandi were Shyan Sen, Gaur Sen, Sigh Se n, Shamsheer Sen, Surma Sen, Ishwari Sen, Jalam Sen, Balbir Sen and Vijay Sen.', 'Many public utility, services and facilities like schools, hospitals and post offices, etc., were started by Raja Vijay Sen during his time.', 'He a lso built large number of palaces and houses.', 'A mule road from Bajjnath in District Kangra to Sultampur in adjoining District of Kullu, via Mandi and the Victoria suspension Bridge on river Beas were constructed during his tin e in the year 1877.', 'Raja Vijay Sen was issue less and left no direct heir.', 'So in 1897, Bhawani Sen was recognized as his successor.', 'During his period the city was considerably damaged in the earthquake of 1905.', 'He was succeeded by Joginder Sen in April 1914 who was, however, vested with full powers of ruling chief in February 1925.', 'During the period of his minority, Mandi remained under the management of a British officer, and a number of departments were overhauled and reorganised.', 'Mandi being one of the oldest city is enjoying urban status since 1981.', 'The Mandi palace popularly known as 'Raj Mahal' as it looked in the past.', 'The area where the 'Sunken']

- Brief description of each segmenter used
 - One or two paragraphs, does it use regular expressions, which programming language

I think both of them (pragmatic segmenter and NLTK Punkt) uses rule based approach for segmentaion. Pragmatic segmeter is written in Ruby whereas NLTK Punkt written in Python.

- Quantitative evaluation: Accuracy percentage (how many sentence boundaries were detected correctly).

Almost all sentence boundaries were detected correctly by both of the segmeters.

- As there are only 10 paragraphs, you can count these manually

There are total 9 paragraphs.

- Qualitative evaluation: What kind of mistakes does each segmenter make?

For this data sample there were one mistake in NLTK Punkt segmenter. It ws not able to recognize this task as two separate sentences, whereas pragmatic segmeter does a good job in this case.

put under the control of Superintendent 'Cis-Satluj States'.Lala Lajpat Rai visited Mandi during 1906

In 1849, Princely states of Mandi, Suket and Chamba were put under the control of Superintendent 'Cis-Satluj States'.Lala Lajpat Rai visited Mandi during 1906 to organize the revolutionary activities.

QUESTIONS from 01b file:

- How should you segment sentences with semi-colon? As a single sentence or as two sentences? Should it depend on context?

Typically, a semicolon is used to connect two closely related clauses [1]. So segmentor can treat it as two separate sentences.

- Should sentences with ellipsis... be treated as a single sentence or as several sentences?
- Ellipsis generally indicate omission of words from quotes, talks or etc. It can be anywhere in the sentence in the middle or at the last. So sentences with ellipsis should be treated a single sentence by identifying them as a single token.

- If there is an exclamation after the first word in the sentence should it be a separate sentence? How about if there is a comma?

It should be treated as a single sentence only as exclamation or comma after first word does not indicate two different sentences.

- Can you think of some hard tasks for the segmenter?

Dealing with abbreviations like U.S.A. or USA can be a hard task for the segmenter.

REFERENCES:

1. <https://writing.wisc.edu/handbook/grammarpunct/semicolons/>