

Learning-Based Named Entity Recognition for Morphologically-Rich, Resource-Scarce Languages

Kazi Saidul Hasan Md. Altaf ur Rahman Vincent Ng

Human Language Technology Research Institute
University of Texas at Dallas

European Chapter of
the Association for Computational Linguistics, 2009

Goal

Named Entity Recognition (NER) for morphologically-rich, resource-scarce languages

- Use **Bengali** as our representative language
- Focus on identifying PERSON, ORGANIZATION, and LOCATION

Challenges

- ❶ Scarcity of NE-labeled data
 - Unsupervised techniques for English NER (e.g., Collins and Singer (1999)) are unlikely to work well
 - **Problem:** Lack of capitalization in Bengali
- ❷ Lack of publicly available gazetteers
- ❸ Inaccurate POS tagger
 - Unsupervised POS induction techniques (e.g., distributional clustering) are unlikely to work well
 - Problems:** Distributional representation may not be reliable for Bengali because of
 - free word order
 - morphological inflections

Weaknesses of Existing Bengali NE Recognizers

- Use their own manually-constructed gazetteers
 - The results are not reproducible
- Use pseudo-affixes (created by extracting the first n and the last n characters of a word)
 - The process is ad-hoc and may not cover many useful affixes
- Typically adopt a **pipelined** NER architecture
 - Errors from the POS tagger are propagated to the NE recognizer

Our Approach

Still supervised, but

- we investigate two new features when used in a pipelined architecture for POS tagging and NER:
 - 1 affixes **induced** from an unannotated corpus
 - 2 semantic class information extracted from Wikipedia
- we propose a **joint** model for learning POS tagging and NER simultaneously

Outline

- 1 Two New Linguistic Features
 - Induced Affixes
 - Semantic Classes from Wikipedia
- 2 POS Tagging with New Features
 - Evaluation
- 3 Pipelined NER with New Features
 - Evaluation
- 4 Joint Model for POS Tagging and NER
 - Evaluation

Outline

- 1 Two New Linguistic Features
 - Induced Affixes
 - Semantic Classes from Wikipedia
- 2 POS Tagging with New Features
 - Evaluation
- 3 Pipelined NER with New Features
 - Evaluation
- 4 Joint Model for POS Tagging and NER
 - Evaluation

Affix Induction

Goal

For morphologically-rich languages, a lot of grammatical information is expressed via affixes. Unlike previous approaches, we learn affixes rather than using pseudo-affixes.

Approach: Keshava and Pitler's (2006) method

Assume that

- (1) V is a vocabulary extracted from a large, unannotated corpus
- (2) α and β are two character sequences
- (3) $\alpha\beta$ is the concatenation of α and β
 - if $\alpha\beta$, $\alpha \in V$, we extract β as a candidate suffix
 - if $\alpha\beta$, $\beta \in V$, we extract α as a candidate prefix

Affix Induction (contd.)

Many of the induced affixes could be **spurious**.

Example: If both “can” and “candidate” are in V , then “didate” is extracted as an induced suffix

Solution

- 1 Assign a score to each induced affix
- 2 Select only those that score over a certain threshold

We use induced affix n-grams as features. These are like word n-grams except that these are made of affixes.

Outline

- 1 Two New Linguistic Features
 - Induced Affixes
 - Semantic Classes from Wikipedia
- 2 POS Tagging with New Features
 - Evaluation
- 3 Pipelined NER with New Features
 - Evaluation
- 4 Joint Model for POS Tagging and NER
 - Evaluation

Semantic Class Induction from Wikipedia

Goal

- 1 Generate a list of phrases and tokens that are potentially named entities from **all** the articles in the Bengali Wikipedia
- 2 Heuristically annotate each of them with one of four classes: PER (person), ORG (organization), LOC (location), or OTHERS

Generating an Annotated List of Phrases

Steps

- 1 Generate and annotate the title of each article
 - (i) Use category information to annotate the title



উত্কলপিডিয়া
একটি মুক্ত বিশ্বকোষ

পরিচয়

- প্রধান পাতা
- বাংলা উইকিপিডিয়া সম্প্রদায়
- সম্প্রদায়
- সম্প্রদায়িক ঘটনা
- সাম্প্রতিক পরিবর্তনসমূহ
- অজানা খেকোনা পৃষ্ঠা
- সাহায্য
- দান করুন

অনুসন্ধান

যাতিয়ার

- সংযোগকারী পৃষ্ঠাসমূহ
- সম্পর্কিত পরিবর্তন
- বিশেষ পৃষ্ঠাসমূহ
- ছাপার বেগা সংস্করণ
- স্থায়ী সংযোগ
- এই নিবন্ধটি উদ্ধৃত করুন

অন্যান্য ভাষাসমূহ

- العربية
- Asturianu
- Azərbaycan
- Žemaitėška
- Беларуская
- Български
- ইমার ঠার/বিশ্বপ্রিয়া
- महिपुरी
- Brezhoneg
- Bosanski

পিথাগোরাস

উইকিপিডিয়া, মুক্ত বিশ্বকোষ থেকে

পিথাগোরাস বা **সামোসের পিথাগোরাস** (প্রাচীন গ্রিক ভাষায় Πυθαγόρας *পুথাগোরাস*) (খ্রিস্টপূর্ব ৫৮০ - খ্রিস্টপূর্ব ৫০০) একজন গ্রীক দার্শনিক ও গণিতবিদ।

পিথাগোরাস সবচেয়ে বেশি বিখ্যাত **পিথাগোরাসের উপপাদ্যের** জন্য। খ্রীষ্টের জন্মের প্রায় ৬০০ বছর আগে পিথাগোরাস দর্শন শাস্ত্রে প্রভূত অবদান রাখেন। তাকে বলা হয় - সংখ্যার জনক। পিথাগোরাস প্রথম নিজেকে দার্শনিক হিসাবে দাবী করেন। *প্লেটো*, *এরিস্টটল* ও কোপারনিকাসের অনেক অর্জন পিথাগোরাসের ধারণায়ই বিস্তৃতি।

তার কোনো লেখা পরে আর পাওয়া যায়নি ফলে তার অবদান সম্পর্কে নিশ্চিত হওয়া যায়নি।

৫২৯ খ্রিস্টপূর্বাব্দে তিনি ইতালীর ক্রোটোনা শহরে বসবাস শুরু করেন।

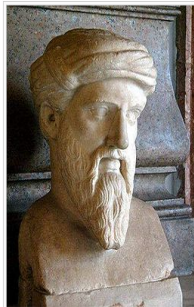
পিথাগোরাসকে বলা যায় সংখ্যা দার্শনিক। তিনি বিশ্বাস করতেন পৃথিবীর সকল কিছুর মূল হলো সংখ্যা। সংখ্যার সঙ্গে নৈতিকতা, রঙ, শব্দ-অশব্দকে তিনি মিরিয়ে ফেলেন। তার বিশ্বাস হলো মানুষের আত্মা অবিনশ্বর। ফলে হিন্দু বৌদ্ধ ও অন্যান্য কয়েকটি ধর্মের মতো তিনিও জন্মান্তরবাদে বিশ্বাসী। পিথাগোরাস পৃথিবীর গোলকত্বে বিশ্বাস করতেন কারণ তার ধারণা ছিল বৃত্ত হলো পরম বক্ররেখা। সূর্য, চন্দ্র বা গ্রহদের নিজেদের একটা গতি আছে।

পিথাগোরাসের উপপাদ্য

[সম্পাদনা]

সমকোণী ত্রিভুজের অতিভুজের ওপর অঙ্কিত বর্গক্ষেত্রের ক্ষেত্রফল অপর দুই বাহুর ওপর অঙ্কিত বর্গক্ষেত্রের ক্ষেত্রফলের সমষ্টির সমান।

এই নিবন্ধটি অসম্পূর্ণ। আপনি চাইলে এটিকে সমৃদ্ধ করতে পারেন।



পিথাগোরাসের প্রতিকৃতি

প্র · আল · স

প্রাক-সক্রেটীয় দার্শনিকবৃন্দ

মিলেভুগীয় ধারা: থেলিস · আনাক্সিমান্ড্রোস · মিলেভুসের আনাক্সিমেনেস

পিথাগোরাসবাদী: পিথাগোরাস · ফিলাপাউস · আঙ্ডামোন · অর্পুতাস · তিমোফাস

এফেসোসীয় ধারা: ইরান্দ্রেইডাস — **এলেশার দার্শনিক ধারা:** স্ক্রেনোকলিড · প্যামেনিডেস · এসেলার জিনো · মেগিলোস

বহরবাদী ধারা: আনাক্সাগোরাস · এস্পেডোক্রেস — **পরমামুবাদী ধারা:** লেভক্সিপ্পাস · দেমোক্রিডোস

সোফিবিদ: প্রোডায়োরাস · গোর্গিয়াস · প্রোদিকোস · ইন্দিবাস

দিওগেনেস · ফেরেকুদেস

বিষয়শ্রেণীসমূহ: অসম্পূর্ণ | গ্রিক গণিতবিদ | গ্রিক দার্শনিক | খ্রিস্টপূর্ব ৫৮০-এ জন্ম | খ্রিস্টপূর্ব ৫০০-এ মৃত্যু



উত্কলিকিভাষা
একটি মুক্ত বিপ্লব

পরিচয়

- প্রধান পাতা
- বাংলা উইকিপিডিয়া সম্প্রদায়
- সমসাময়িক ঘটনা
- সাম্প্রতিক পরিবর্তনসমূহ
- অজানা খেকোনা পৃষ্ঠা
- সাহায্য
- দান করুন

অনুসন্ধান

যাতিয়ার

- সংযোগকারী পৃষ্ঠাসমূহ
- সম্পর্কিত পরিবর্তন
- বিশেষ পৃষ্ঠাসমূহ
- ছাপার বেগা সংস্করণ
- স্থায়ী সংযোগ
- এই নিবন্ধটি উদ্ধৃত করুন

অন্যান্য ভাষাসমূহ

- العربية
- Asturianu
- Azərbaycan
- Žemaitėška
- Беларуская
- Български
- ইমার ঠার/বিশ্বপ্রিয়া
- मगधुनी
- Brezhoneg
- Bosanski

লেখক: অ্যাভাক্স: সশাংলি: ক্রসল: খতখস:

How to read and write in Unicode Bangla Must-read for contributors

পিথাগোরাস

Pythagoras

উইকিপিডিয়া, মুক্ত বিশ্বকোষ

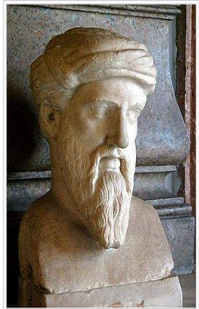
পিথাগোরাস বা **সামোসের পিথাগোরাস** (প্রাচীন গ্রিক ভাষায় Πυθαγόρας *পুথাগোরাস*) (খ্রিস্টপূর্ব ৫৮০ - খ্রিস্টপূর্ব ৫০০) একজন গ্রীক দার্শনিক ও গণিতবিদ।

পিথাগোরাস সবচেয়ে বেশি বিখ্যাত **পিথাগোরাসের উপপাদ্যের** জন্য। খ্রীষ্টের জন্মের প্রায় ৬০০ বছর আগে পিথাগোরাস দর্শন শাস্ত্রে প্রভূত অবদান রাখেন। তাকে বলা হয় - সংখ্যার জনক। পিথাগোরাস প্রথম নিজেকে দার্শনিক হিসাবে দাবী করেন। *প্লেটো*, *এরিস্টটল* ও কোপারনিকাসের অনেক অর্জন পিথাগোরাসের ধারণায়ই বিস্তৃতি।

তার কোনো লেখা পরে আর পাওয়া যায়নি ফলে তার অবদান সম্পর্কে নিশ্চিত হওয়া যায়নি।

৫২৯ খ্রিস্টপূর্বাব্দে তিনি ইতালীর ক্রোটোনা শহরে বসবাস শুরু করেন।

পিথাগোরাসকে বলা যায় সংখ্যা দার্শনিক। তিনি বিশ্বাস করতেন পৃথিবীর সকল কিছুর মূল হলো সংখ্যা। সংখ্যার সঙ্গে নৈতিকতা, রঙ, শব্দ-অশব্দকে তিনি মিরিয়ে ফেলেন। তার বিশ্বাস হলো মানুষের আত্মা অবিনশ্বর। ফলে হিন্দু বৌদ্ধ ও অন্যান্য কয়েকটি ধর্মের মতো তিনিও জন্মান্তরবাদে বিশ্বাসী। পিথাগোরাস পৃথিবীর গোলকত্বে বিশ্বাস করতেন কারণ তার ধারণা ছিল বৃত্ত হলো পরম বক্ররেখা। সূর্য, চন্দ্র বা গ্রহদের নিজেদের একটা গতি আছে।



পিথাগোরাসের প্রতিকৃতি

পিথাগোরাসের উপপাদ্য

[সম্পাদনা]

সমকোণী ত্রিভুজের অতিভুজের ওপর অঙ্কিত বর্গক্ষেত্রের ক্ষেত্রফল অপর দুই বাহুর ওপর অঙ্কিত বর্গক্ষেত্রের ক্ষেত্রফলের সমষ্টির সমান।

এই নিবন্ধটি অসম্পূর্ণ। আপনি চাইলে এটিকে সমৃদ্ধ করতে পারেন।

প্রাক-সক্রেটীয় দার্শনিকবৃন্দ

মিলেভুসীয় ধারা: থেলিস • আনাক্সিমান্ড্রোস • মিলেভুসের আনাক্সিমেদিস

পিথাগোরাসবাদী: পিথাগোরাস • ফিলাপাউস • আঙ্ডমাবোন • অর্পুতাস • তিমাথোস

এফেসোসীয় ধারা: ইয়ার্রেইডাস — **এলেশার দার্শনিক ধারা:** স্ক্রেনোকলিজ • প্যামেনিডেস • এসেলার জিনো • মেগিলোস

বহরবাদী ধারা: আনাক্সাগোরাস • এস্পেডোক্রেস — **পরমাসুবাদী ধারা:** লেভক্সিপ্রাস • দেমোক্রিডোস

সোফিবিদ: প্রোডায়োরাস • গোর্গিয়াস • প্রোদিকোস • ইন্দিবাস

দিওগেনেস • ফেরেকুদেস

বিষয়শ্রেণীসমূহ: অসম্পূর্ণ • গ্রিক গণিতবিদ • গ্রিক দার্শনিক | **খ্রিস্টপূর্ব ৫৮০-এ জন্ম** | **খ্রিস্টপূর্ব ৫০০-এ মৃত্যু**



ଉତ୍କଳିକାପିଡ଼ିଆ
ଏକটি ମୁକ୍ତ ବିଶ୍ୱକୋଷ

ପରିଚୟ

- ପ୍ରଧାନ ପାଠା
- ବାହା ଉଇକିପିଡ଼ିଆ
- ସମ୍ପ୍ରଦାୟ
- ସମ୍ପାଦନା
- ସାମ୍ପ୍ରତିକ ପରିବର୍ତ୍ତନସମୂହ
- ଅଜ୍ଞାନା ଖୋଜାଣା ପୃଷ୍ଠା
- ସାହାଯ୍ୟ
- ଦାନ କରନ୍

ଅନୁସନ୍ଧାନ

ସାହାଯ୍ୟ

- ସାମ୍ପ୍ରତିକ ପରିବର୍ତ୍ତନସମୂହ
- ସମ୍ପ୍ରତିକ ପରିବର୍ତ୍ତନ
- ବିଶେଷ ପୃଷ୍ଠାସମୂହ
- ଘୋଷଣା ସାହାଯ୍ୟ
- ସୂଚୀ ସାମ୍ପ୍ରତିକ
- ଏହି ନିବନ୍ଧଟି ଉଦ୍ଧୃତ କରନ୍

ଅନ୍ୟାନ୍ୟ ଭାଷାସମୂହ

- العربية
- Asturianu
- Azərbaycan
- Žemaitėška
- Беларуская
- Български
- ཇོ་མོ་གླང་མ་
- ཇོ་མོ་གླང་མ་
- Brezhoneg
- Bosanski

How to read and write in Unicode Bangla Must-read for contributors

ପିথাଗୋରାସ

Pythagoras

ଉଇକିପିଡ଼ିଆ, ମୁକ୍ତ ବିଶ୍ୱକୋଷ

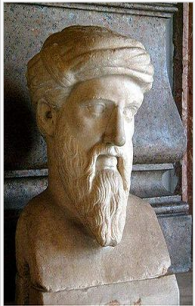
ପିଥାଗୋରାସ ବା **ସାମୋସର ପିଥାଗୋରାସ** (ପ୍ରାଚୀନ ଗ୍ରିକ ଭାଷା Pythagorás *ପୁଥାଗୋରାସ*) (**ଖ୍ରୀଷ୍ଟପୂର୍ବ ୫୮୦ - ଖ୍ରୀଷ୍ଟପୂର୍ବ ୫୦୦**) ଏକଜନ ଗ୍ରିକ ଦାର୍ଶନିକ ଓ ଗଣିତବିଦ।

ପିଥାଗୋରାସ ସବୁଜଣେ ବେଶି ବିখ୍ୟାତ **ପିଥାଗୋରାସର ଉପପାଦର** ଜ୍ଞା। ଖ୍ରୀଷ୍ଟର ଜନ୍ମର ପ୍ରାୟ ୬୦୦ ବର୍ଷ ଆଗେ ପିଥାଗୋରାସ ଦର୍ଶନ ଶାସ୍ତ୍ର ପ୍ରଭୃତ ଅବଦାନ ରାଖିଲେ। ତାଙ୍କ ବଂଶ ହେଉଛି - ସଂଖ୍ୟାର ଜନକ। ପିଥାଗୋରାସ ପ୍ରଥମ ନିଜେ ଦାର୍ଶନିକ ହିସାବେ ଦାବୀ କଲେ। **ମେଟା, ଏରିଷ୍ଟଟଲ** ଓ କୋପାରନିକାସର ଅନେକ ଅର୍ତ୍ତନ ପିଥାଗୋରାସର ଧାରଣାରେ ବିସ୍ତୃତ।

ତାର କୋନୋ ଲେଖା ପରେ ଆଉ ପାଠ୍ୟା ଯାହା ଧରି ତାର ଅବଦାନ ସମ୍ପର୍କେ ନିଶ୍ଚିତ ହେଉ ଯାହା।

୧୨୦ ଖ୍ରୀଷ୍ଟପୂର୍ବାଦେ ତିନି ଇଟାଲୀର କ୍ରେଟୋନା ନ୍ଦରେ ବସବାସ କଲେ।

ପିଥାଗୋରାସକେ ବନା ଯାଏ ସଂଖ୍ୟା ଦାର୍ଶନିକ। ତିନି ବିଶ୍ୱାସ କଲେନ ପୃଥିବୀର ସକଳ କିନ୍ତୁର ମୂଳ ହେଉ ସଂଖ୍ୟା। ସଂଖ୍ୟାର ସତ୍ତ୍ୱେ ନୈତିକତା, ଋଷ୍ଟ, ଶୁଦ୍ଧ-ଅଶୁଦ୍ଧକେ ତିନି ଧରିବେ କଲେନ। ତାର ବିଶ୍ୱାସ ହେଉ ଯେ ମାନୁଷ୍ୟର ଆତ୍ମା ଅବିନଶ୍ୱର। ତାହା ହିନ୍ଦୁ ବୌଦ୍ଧ ଓ ଅନ୍ୟାନ୍ୟ କାହାକି ଧର୍ମର ମତେ ତିନି ଓ ଜନ୍ମାନ୍ତରବାଦେ ବିଶ୍ୱାସୀ। ପିଥାଗୋରାସ ପୃଥିବୀର ଗୋଳକ୍ରେ ବିଶ୍ୱାସ କଲେନ କାରଣ ତାର ଧାରଣା ଦିନ ବୃତ୍ତ ହେଉ ପରମ ବସ୍ତୁତା। ସୂର୍ଯ୍ୟ, ଚନ୍ଦ୍ର ବା ଗ୍ରହଣର ନିଜେନେର ଏକଟା ଗତି ହେଉ।



ପିଥାଗୋରାସର ପ୍ରତିମା

ପିଥାଗୋରାସର ଉପପାଦ

[ସମ୍ପାଦନା]

ସମକାଳୀ ତ୍ରିଭୁଜର ଅତିଭୁଜର ଓପର ଅଞ୍ଚିତ ବର୍ଗକ୍ଷେତ୍ରର କ୍ଷେତ୍ରଫଳ ଅପର ଦୁଇ ବାହୁର ଓପର ଅଞ୍ଚିତ ବର୍ଗକ୍ଷେତ୍ରର କ୍ଷେତ୍ରଫଳର ସମାନ।

ଏହି ନିବନ୍ଧଟି ଅସମ୍ପୂର୍ଣ୍ଣ। ଆପଣ ଯଦି ଏହାକୁ ସମ୍ପୂର୍ଣ୍ଣ କରନ୍, ତେବେ ଆପଣ ଏହାକୁ ସମ୍ପୂର୍ଣ୍ଣ କରିପାରିବେ।

ପ୍ରାକ-ସଂସ୍କୃତି ଦାର୍ଶନିକବୃନ୍ଦ

Categories: Greek mathematician & philosopher |
Born: 580 BC | Death: 500 BC

ନୋ • ମେମ୍ବର
ମତାମତ

ବିଷୟାବଳୀ: ଅସମ୍ପୂର୍ଣ୍ଣ | ଗ୍ରିକ ଗଣିତବିଦ | ଗ୍ରିକ ଦାର୍ଶନିକ | ଖ୍ରୀଷ୍ଟପୂର୍ବ ୫୮୦-ଏ ଜନ୍ମ | ଖ୍ରୀଷ୍ଟପୂର୍ବ ୫୦୦-ଏ ମୃତ୍ୟୁ



উইকিপিডিয়া
একটি মুক্ত বিশ্বকোষ

পরিভ্রমণ

- প্রধান পাতা
- বাংলা উইকিপিডিয়া সম্প্রদায়
- সম্পাদক
- সম্প্রদায়িক ঘটনা
- সাম্প্রতিক পরিবর্তনসমূহ
- অজানা খেকোনা পৃষ্ঠা
- সহায়তা
- দান করুন

অনুসন্ধান

যাতিয়ার

- সংযোগকারী পৃষ্ঠাসমূহ
- সম্পর্কিত পরিবর্তন
- বিশেষ পৃষ্ঠাসমূহ
- ছাপার বোনা সংস্করণ
- স্থায়ী সংযোগ
- এই নিবন্ধটি উদ্ধৃত করুন

অন্যান্য ভাষাসমূহ

- العربية
- Asturianu
- Azərbaycan
- Žemaitėška
- Беларуская
- Български
- ইমার ঠার/বিশ্বপ্রিয়া
- महिपुरी
- Brezhoneg
- Bosanski

উইকিপিডিয়া: সূচীপত্র | সম্পাদনা | ফর্ম | খবর

How to read and write in Unicode Bangla Must-read for contributors

পিথাগোরাস

উইকিপিডিয়া, মুক্ত বিশ্বকোষ

Pythagoras

পিথাগোরাস বা **সামোসের পিথাগোরাস** (প্রাচীন গ্রিক ভাষায় Πυθαγόρας *পুথাগোরাস*) (খ্রিস্টপূর্ব ৫৮০ - খ্রিস্টপূর্ব ৫০০) একজন গ্রীক দার্শনিক ও গণিতবিদ।

পিথাগোরাস সবচেয়ে বেশি বিখ্যাত **পিথাগোরাসের উপপাদ্যের** জন্য। খ্রীষ্টের জন্মের প্রায় ৬০০ বছর আগে পিথাগোরাস দর্শন শাস্ত্রে প্রভূত অবদান রাখেন। তাকে বলা হয় - সংখ্যার জনক। পিথাগোরাস প্রথম নিজেকে দার্শনিক হিসাবে দাবী করেন। **প্লেটো**, **এরিস্টটল** ও কোপারনিকাসের অনেক অর্জন পিথাগোরাসের ধারণারই বিস্তৃতি।

তার কোনো লেখা পরে আর পাওয়া যায়নি ফলে তার অবদান সম্পর্কে নিশ্চিত হওয়া যায়নি।

৫২৯ খ্রিস্টপূর্বাব্দে তিনি ইতালীর ক্রোটোনা শহরে বসবাস শুরু করেন।

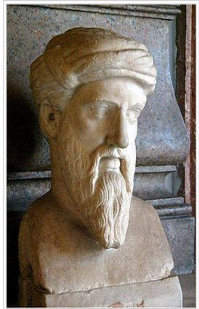
পিথাগোরাসকে বলা যায় সংখ্যা দার্শনিক। তিনি বিশ্বাস করতেন পৃথিবীর সকল কিছুর মূল হলো সংখ্যা। সংখ্যার সঙ্গে নৈতিকতা, রঙ, শব্দ-অশব্দকে তিনি মিশিয়ে ফেলেন। তার বিশ্বাস হলো মানুষের আত্মা অবিনশ্বর। সপ্ত দিন বৌদ্ধ ও অন্যান্য কাহিনী মানে মতো তিনিও জন্মান্তরবাদে বিশ্বাসী। পিথাগোরাস পৃথিবীর গোলকত্ব বিশ্বাস করতেন কারণ তার নিজাদের একটা গতি আছে।

PER	Born on, Death on
LOC	Cities of, Countries of

পিথাগোরাসের উপপাদ্য

সমকোণী ত্রিভুজের অতিভুজের ওপর অঙ্কিত বর্গক্ষেত্রের ক্ষেত্রফল অপর দুই বাহুর

এই নিবন্ধটি অসম্পূর্ণ। আপনি চাইলে এটিকে সম্পূর্ণ করতে পারেন।



পিথাগোরাসের প্রতিকৃতি

প্র - অ - র

Categories: Greek mathematician & philosopher |
Born: 580 BC | Death: 500 BC

নো • মেনিসোস
মডেল

বিশ্বশ্রেণীসমূহ: অসম্পূর্ণ | গ্রিক গণিতবিদ | গ্রিক দার্শনিক | খ্রিস্টপূর্ব ৫৮০-এ জন্ম | খ্রিস্টপূর্ব ৫০০-এ মৃত্যু

Generating an Annotated List of Phrases (contd.)

Steps

- 1 Generate and annotate the title of each article
 - (i) Use category information to annotate the title
 - (ii) Use a small set of keywords to annotate the title

Generating an Annotated List of Phrases (contd.)

NE Class	Keywords
PER	“born,” “died,” “one,” “famous”
LOC	“city,” “area,” “population,” “located,” “part of”
ORG	“establish,” “situate,” “publish”

Table: Keywords for each named entity class



উইকিপিডিয়া
একটি মুক্ত বিশ্বকোষ

নিবন্ধ

আলোচনা

সম্পাদনা করুন

ইতিহাস

 **How to read and write in Unicode Bangla**  **Must-read for contributors**

স্থানাঙ্ক: ২২.৫৬৭১, ৮৮.৩৬৭১

কলকাতা

উইকিপিডিয়া, মুক্ত বিশ্বকোষ থেকে

কলকাতা (ইংরেজি ভাষায়: Kolkata, পূর্বে Calcutta^{সহিতকামত}) ভারতের পূর্ববঙ্গীয় রাজ্য পশ্চিমবঙ্গের রাজধানী। হুগলী নদীর পূর্ব তীরে অবস্থিত এই শহরের পৌর অঞ্চলের জনসংখ্যা ৫০ লক্ষের কিছু বেশি। তবে কলকাতা ও তার পার্শ্ববর্তী জেলাগুলিতে বিস্তৃত কলকাতার মহানগরীয় অঞ্চলের জনসংখ্যা ১ কোটি ৪০ লক্ষের কাছাকাছি; এই জনসংখ্যার বিচারে কলকাতা ভারতের চতুর্থ বৃহত্তম শহর ও তৃতীয় বৃহত্তম মেট্রোপলিটান বা মহানগরীয় অঞ্চল।

১৯১১ সাল পর্যন্ত কলকাতা ব্রিটিশ ভারতের রাজধানী ছিল। একদা ভারতের আধুনিক শিক্ষা, বিজ্ঞান, শিল্প, সংস্কৃতি ও রাজনীতির কেন্দ্রভূমি কলকাতা মহানগরে ১৯৫৪ সালের পর থেকে তীব্র রাজনৈতিক সংঘাত ও তার সঙ্গে সঙ্গে অর্থনৈতিক মন্দা দেখা দেয়। তবে ২০০০ সালের পর থেকে এই শহর আবার আর্থিক ও বাণিজ্যিক সমৃদ্ধির পথে অগ্রসর হয় ও সাম্প্রতিক হজগৌরব অনেকাংশে পুনরাধিকার করে। তবে ভারতের অন্যান্য মহানগরগুলির মতো নগরায়নজনিত দারিদ্র্য, পরিবেশ দূষণ ও বানভাটের সমস্যা থেকে কলকাতাও একবারে মুক্ত হতে পারেনি।

কলকাতা শহরের প্রসিদ্ধি তার বৈশ্বিক আন্দোলনগুলির জন্য। ভারতের স্বাধীনতা আন্দোলন এবং পরবর্তীতে বামপন্থী বাণিজ্যিক ইউনিয়নগুলির আন্দোলন এ শহরের ইতিহাসের একটি বড় অংশ। এর সঙ্গে সঙ্গে আধুনিক ভারতে কলকাতা *সাম্প্রতিক রাজধানী* ও *আনন্দ নগরী* (City of Joy সিটি অব জয়) নামে নথিত। রবীন্দ্রনাথ ঠাকুর, রোনাল্ড রস, সুভাষচন্দ্র বসু, মাদার তেরেসা, সভ্যজিৎ রায়, সত্যেন্দ্রনাথ বসু, স্বামী বিবেকানন্দ রাজা রামমোহন রায় সি ভি রামন-সহ বহু বিশ্ববিদ্রুত ব্যক্তিরের বাসভূমি এই নগরী তার ঐতিহাসিক ঐতিহ্য ও সমৃদ্ধ সাম্প্রতিক প্রেক্ষাপটের কারণে আজও বিশ্ববাসীর চোখে মর্যাদার আসনে অধিষ্ঠিত।

সূত্রপত্র [আড়ালে রাখা]

- নামকরণ
- ইতিহাস
- ভূগোল
 - শহরের গঠন
- অর্থনীতি
- রাজধানী ও নগর প্রশাসন
- বাতায়োত
- জনপরিসংখ্যান
- সংস্কৃতি
- শিক্ষা
- গণমাধ্যম

কলকাতা

পশ্চিমবঙ্গ • ভারত



ভিক্টোরিয়া মেমোরিয়াল



স্থানাঙ্ক: ২২.৫৬৭১, ৮৮.৩৬৭১

সময় IST (UTC+৫:৩০)

অঞ্চল

অবস্থান ১,৫০০ বর্গ কিমি (৫৭৯ sq mi)

• উচ্চতা • ১০ m (৩০ ft)

দেশ কলকাতা জেলা

জনসংখ্যা ৪,৫৬০,৫৪৪ (২০০১)

• ঘনত্ব

• ঘনত্ব ২৯,২৬১/কিমি^২ (৭৫,৫০০/বর্গ মাইল) (এক্সপ্রেশন ত্রুটি: <=-এর জন্য অপেক্ষারত নেই।

পরিসর

- প্রধান পাতা
- বাংলা উইকিপিডিয়া সম্প্রদায়
- সমসাময়িক ঘটনা
- সাম্প্রতিক পরিবর্তনসমূহ
- অজানা বোকানো পৃষ্ঠা
- সাহায্য
- দান করুন

অনুসন্ধান

চলো

অনুসন্ধান

হাতিয়ার

- সংযোগকারী পৃষ্ঠাসমূহ
- সম্পর্কিত পরিবর্তন
- বিশেষ পৃষ্ঠাসমূহ
- ছাপার বোগা সংস্করণ
- স্থায়ী সংযোগ
- এই নিবন্ধটি উদ্ধৃত করুন

অন্যান্য ভাষাসমূহ

- Aragonés
- العربية
- অসমীয়া
- Azərbaycan
- Беларуская
- Български
- বাংলা
- Brezhoneg



উইকিপিডিয়া
একটি মুক্ত বিশ্বকোষ

নিবন্ধ

আলোচনা

সম্পাদনা করুন

ইতিহাস

How to read and write in Unicode Bangla

Must-read for contributors

কলকাতা

Kolkata

স্থানাঙ্ক: ২২.৫৬৭১, ৮৮.৩৬৭১

উৎসপিণ্ডিবা, মুক্ত বিশ্বকোষ থেকে

কলকাতা (ইংরেজি ভাষায়: Kolkata, পূর্বে Calcutta ^{সহিত্যকৃত}) ভারতের পূর্ববঙ্গীয় রাজ্য পশ্চিমবঙ্গের রাজধানী। হুগলী নদীর পূর্ব তীরে অবস্থিত এই শহরের পৌর অঞ্চলের জনসংখ্যা ৫০ লক্ষের কিছু বেশি। তবে কলকাতা ও তার পার্শ্ববর্তী জেলাগুলিতে বিস্তৃত কলকাতার মহানগরীয় অঞ্চলের জনসংখ্যা ১ কোটি ৪০ লক্ষের কাছাকাছি; এই জনসংখ্যার বিচারে কলকাতা ভারতের চতুর্থ বৃহত্তম শহর ও তৃতীয় বৃহত্তম মেট্রোপলিটান বা মহানগরীয় অঞ্চল।

১৯১১ সাল পর্যন্ত কলকাতা ব্রিটিশ ভারতের রাজধানী ছিল। একদা ভারতের আধুনিক শিক্ষা, বিজ্ঞান, শিল্প, সংস্কৃতি ও রাজনীতির কেন্দ্রভূমি কলকাতা মহানগরে ১৯৫৪ সালের পর থেকে তীর্থ রাজনৈতিক সংঘাত ও তার সঙ্গে সঙ্গে অর্থনৈতিক মন্দা দেখা দেয়। তবে ২০০০ সালের পর থেকে এই শহর আবার আর্থিক ও বাণিজ্যিক সমৃদ্ধির পথে অগ্রসর হয় ও সাংস্কৃতিক হজগৌরব অনেকাংশে পুনরাধিকার করে। তবে ভারতের অন্যান্য মহানগরগুলির মতো নগরায়নজনিত দারিদ্র্য, পরিবেশ দূষণ ও যানজটের সমস্যা থেকে কলকাতাও একবারে মুক্ত হতে পারেনি।

কলকাতা শহরের প্রসিদ্ধি তার বৈশ্বিক আন্দোলনগুলির জন্য। ভারতের স্বাধীনতা আন্দোলন এবং পরবর্তীতে বামপন্থী বাণিজ্যিক ইউনিয়নগুলির আন্দোলন এ শহরের ইতিহাসের একটি বড় অংশ। এর সঙ্গে সঙ্গে আধুনিক ভারতে কলকাতা *সাংস্কৃতিক রাজধানী* ও *আনন্দ নগরী* (City of Joy সিটি অব জয়) নামে নথিত। রবীন্দ্রনাথ ঠাকুর, রোনাল্ড রস, সুভাষচন্দ্র বসু, মাদার তেরেসা, সভ্যজিৎ রায়, সত্যেন্দ্রনাথ বসু, স্বামী বিবেকানন্দ রাজা রামমোহন রায় সি ভি রাসন-সহ বহু বিখ্যাত ব্যক্তির বাসভূমি এই নগরী তার ঐতিহাসিক ঐতিহ্য ও সমৃদ্ধ সাংস্কৃতিক প্রেক্ষাপটের কারণে আজও বিশ্ববাসীর চোখে মর্যাদার আসনে অধিষ্ঠিত।

সূচিপত্র [আড়ালে রাখা]

- নামকরণ
- ইতিহাস
- ভূগোল
 - শহরের গঠন
- অর্থনীতি
- রাজধানী ও নগর প্রশাসন
- যাতায়াত
- জনপরিসংখ্যান
- সংস্কৃতি
- শিক্ষা
- গণমাধ্যম

কলকাতা

পশ্চিমবঙ্গ • ভারত



ভিক্টোরিয়া মেমোরিয়াল



স্থানাঙ্ক: ২২.৫৬৭১, ৮৮.৩৬৭১

সময় IST (UTC+৫:৩০)

অঞ্চল

অবস্থান ১,৫০০ বর্গ কিমি (৫৭৯ sq mi)

• উচ্চতা • ৯ m (৩০ ft)

দেশ কলকাতা জেলা

জনসংখ্যা ৪,৫৮০,৫৪৪ (২০০১)

• ঘনত্ব •

• ভাষা(সমূহ) (একপ্রধান ভাষা): বাংলা, অসমীয়া, হিন্দি, ইংরেজি

উইকিপিডিয়া
একটি মুক্ত বিশ্বকোষ

ପରିଚ୍ଛେଦନ

- প্রধান পাতা
- বাংলা উইকিপিডিয়া সম্প্রদায়
- সমসাময়িক ঘটনা
- সাম্প্রতিক পরিবর্তনসমূহ
- অজানা যেকোনো পৃষ্ঠা
- সাহায্য
- দান করুন

অনুসন্ধান

চলো

অনুসন্ধান

হাতিয়ার

- সংযোগকারী পৃষ্ঠাসমূহ
- সম্পর্কিত পরিবর্তন
- বিশেষ পৃষ্ঠাসমূহ
- ছাপার ব্যোগ্য সংস্করণ
- স্থায়ী সংযোগ
- এই নিবন্ধটি উদ্ধৃত
ককন

অন্যান্য ভাষাসমূহ

- Aragonès
- العربية
- অসমীয়া
- Azərbaycan
- Беларуская
- Български
- བོད་སྐད་
- Brezhoneg

निबन्ध

আলোচনা

सम्पादना करण

ইতিহাস

 [How to read and write in Unicode Bangla](#)

 **Must-read for contributors**

କଳକାତା

Kolkata

উৎকলসিঙিয়া, মক্ত বিশ্বকোষ থেকে

কলকাতা (ইংরেজি ভাষায়: Kolkata, পূর্বে **Calcutta** সার্বভৌমত্ব) ভারতের পূর্বাঞ্চলীয় রাজ্য পশ্চিমবঙ্গের রাজধানী। হলদী নদীর পূর্ব তীরে অবস্থিত এই শহরের পৌর অঞ্চলের জনসংখ্যা ৫০ লক্ষের কিছু বেশি। তবে কলকাতা ও তার পার্শ্ববর্তী কাছাকাছি: এই জনসংখ্যার মহানগরীয় অঞ্চল।

located রাজ্যের মা কোটি ৪০ লক্ষের
র র চতুর্ মনোপলিত বা

population

১৯৬৮ তা **ব্রিটিশ ভারতের** রাজধানী ছিল। একদা ভারতের আধুনিক শিক্ষা, বিজ্ঞান, শিল্প, সংস্কৃতি ও রাষ্ট্রনীতির পটভূমি গড়ে তুলেছিলেন এই রাজধানীতে বসে। **area** কলকাতা মহানগরে **১৯৫৪** সালের পর থেকে জরুরী রাজনৈতিক সংঘাত ও তার সঙ্গে সঙ্গে অর্থনীতির অবনতি ঘটেছিল। তবে **২০০০** সালের পর থেকে এই শহর আবার আর্থিক ও বাণিজ্যিক সমৃদ্ধির পথে অগ্রসর হয় ও সামুদ্রিক হস্তশিল্পের অনেকাংশই পুনরাবিষ্কার করে। তবে ভারতের অন্যান্য মহানগরগুলোর মতো নারায়ণগঞ্জও দরিদ্রতা, পরিবেশ দূষণ ও যানজটের সমস্যা থেকে কলকাতা ও একাধারে মুক্ত থাকেনি।

কলকাতা শহরের প্রসিদ্ধি তার বৈশিষ্ট্য আন্দোলনগুলির জন্য। ভারতের স্বাধীনতা আন্দোলন এবং পরবর্তীতে বামপন্থী আন্দোলন এ শহরের ঐতিহ্যমণ্ডিত একটি বড় অংশ। এর সঙ্গে সঙ্গে আধুনিক ভারতে famous শহরী ও আনন্দ নগরী (City of Joy) নামেও পরিচিত। রবীন্দ্রনাথ ঠাকুর, বসু, মান্নার ভেলেসা, সত্যজিৎ রায়, part of বিবেকানন্দ রাজা রামসাহেন রায় সি ভি রামেন-সহ বহু বিখ্যাত ব্যক্তিদের বাসভূমি। এছাড়াও প্রাইভেট ও সন্মুখ সাংস্কৃতিক প্রেক্ষাপটের কারণে আজও বিশ্ববাসীরা চোখে মগ্নমান আনন্দ আনন্দ।

সৃষ্টিপত্র [আডালে রাখো]

- ১ নামকরণ
- ২ ইতিহাস
- ৩ ভূগোল
- ৩.১ শহরের গঠন
- ৪ অর্থনীতি
- ৫ রাজধানী ও নগর প্রশাসন
- ৬ যাতায়াত
- ৭ জনসংস্কৃতি
- ৮ সংস্কৃতি
- ৯ শিক্ষা
- ১০ গণমাধ্যম

କଳକାଞ୍ଚା

शुद्धिपत्र : आचार्य



ভিস্টে'রিয়া মেমোরিয়াল



शालाह: 22.5697, 88.3697

प्रत्यक्ष IST (UTC+५:३०)

अज्ञान

আয়তন ১,৫০০ বর্গ কিমি (৫৭৯ sq mi)

- ઉંચાઈ • ૨ m (૭૦ ft)

জেলা কলকাতা জেলা

.....

जनसंख्या

৯.১২০৮ (একপ্রশ্ন-কৃতি: ৫-এর জন্য অর্ধসংকেত নাই) ০



উইকিপিডিয়া
একটি মুক্ত বিশ্বকোষ

নিবন্ধ

আলোচনা

সম্পাদনা করুন

ইতিহাস

How to read and write in Unicode Bangla

Must-read for contributors

কলকাতা

Kolkata

স্থানাঙ্ক: ২২.৫৬৭১, ৮৮.৩৬৭১

উৎসপিডিয়া, মুক্ত বিশ্বকোষ থেকে

কলকাতা (ইংরেজি ভাষায়: Kolkata, পূর্বে Calcutta ^{সহচর্য}) ভারতের পূর্বাঞ্চলীয় রাজ্য পশ্চিমবঙ্গের রাজধানী। হুগলী নদীর পূর্ব তীরে অবস্থিত এই শহরের পৌর অঞ্চলের জনসংখ্যা ৫০ লক্ষের কিছু বেশি। কলকাতা ও তার পার্শ্ববর্তী এলাকায় কলকাতার মহানগরী নামে একটি মেট্রো এলাকা রয়েছে। কাছাকাছি; এই জনসংখ্যার কারণে এলাকাটিকে ভারতের চতুর্থ বৃহত্তম মেট্রো এলাকা হিসেবে গণ্য করা হয়।

weight = 3

located

population

area

১৯১১ সালে ব্রিটিশ ভারতের রাজধানী ছিল। একদা ভারতের আধুনিক শিক্ষা, বিজ্ঞান, শিল্প, সংস্কৃতি ও রাষ্ট্রনীতির কেন্দ্রবিন্দু হিসেবে কলকাতা মহানগরে ১৯৫৪ সালের পর থেকে তীব্র রাজনৈতিক সংঘাত ও তার সঙ্গে অর্থনৈতিক দুর্য্যোগের কারণে কলকাতার গুরুত্ব হ্রাস পেয়েছে। তবে ২০০০ সালের পর থেকে এই শহর আবার আর্থিক ও বাণিজ্যিক সমৃদ্ধির পথে অগ্রসর হয় ও সামাজিক সহজগৌরব অনেকাংশে পুনরাধিকার করে। তবে ভারতের অন্যান্য মহানগরগুলির মতো নগরায়নজনিত দারিদ্র্য, পরিবেশ দূষণ ও বানভাঙের সমস্যা থেকে কলকাতাও একবারে মুক্ত হতে পারেনি।

কলকাতা শহরের প্রসিদ্ধি তার বৈশিষ্ট্যবিশিষ্ট আন্দোলনগুলির জন্য। ভারতের স্বাধীনতা আন্দোলন এবং পশ্চিমবঙ্গের স্বাধীনতা আন্দোলন এ শহরের ইতিহাসের একটি বড় অংশ। এর সঙ্গে সঙ্গে ভারতের স্বাধীনতা আন্দোলন ও আন্দোলন নগরী (City of Joy) নামেও পরিচিত। শহরটি ঠাকুর, বসু, মাদার তেরেসা, সভ্যজিৎ রায়, রামসাহেন রায় সি ভি রাসন-সহ বহু বিখ্যাত ব্যক্তিদের বাসভূমি। সামাজিক প্রেক্ষাপটের কারণে আজও বিমুখবাসীর চোখে মর্মান্বণ আসনে আশ্রিত।

weight = 1

famous

part of

সৃষ্টিকর্তা [আড়ালো রাখা]

- ১ নামকরণ
- ২ ইতিহাস
- ৩ ভূগোল
 - ৩.১ শহরের গঠন
- ৪ অর্থনীতি
- ৫ রাজধানী ও নগর প্রশাসন
- ৬ বাতায়ন
- ৭ জনপরিসংখ্যান
- ৮ সংস্কৃতি
- ৯ শিক্ষা
- ১০ গণমাধ্যম

কলকাতা

পশ্চিমবঙ্গ • ভারত



ভিক্টোরিয়া মেমোরিয়াল



স্থানাঙ্ক: ২২.৫৬৭১, ৮৮.৩৬৭১

সময়	IST (UTC+৫:৩০)
অঞ্চল	
অবস্থান	১,৫০০ বর্গ কিমি (৫৭৯ sq mi)
উচ্চতা	২ m (৬০ ft)
দেশ	কলকাতা, বেঙ্গা
জনসংখ্যা	৪,৫৮০,৫৪৪ (২০০১)
ভাষা	• বাংলা
সময়	২০২২ (এপ্রিল ২০২২) (এপ্রিল ২০২২) (এপ্রিল ২০২২) (এপ্রিল ২০২২)

- প্রধান পাতা
- বাংলা উইকিপিডিয়া সম্প্রদায়
- সমসাময়িক ঘটনা
- সাপ্তাহিক পরিবর্তনসমূহ
- অজানা কোনো পৃষ্ঠা
- সাহায্য
- দান করুন

অনুসন্ধান

চলো অনুসন্ধান

- হাতিয়ার
- সংযোগকারী পৃষ্ঠাসমূহ
 - সম্পর্কিত পরিবর্তন
 - বিশেষ পৃষ্ঠাসমূহ
 - ছাপার যোগ্য সংস্করণ
 - স্থায়ী সংযোগ
 - এই নিবন্ধটি উন্নত করুন

- অন্যান্য ভাষাসমূহ
- Aragonés
 - العربية
 - অসমীয়া
 - Azərbaycan
 - Беларуская
 - Български
 - Български
 - Brezhoneg



১১৬

- প্রধান পাতা
বাংলা উইকিপিডিয়া
সম্প্রদায়
সমসাময়িক ঘটনা
সাম্প্রতিক পরিবর্তনসমূহ
অজানা যেকোনো পৃষ্ঠা
সাহায্য
দান করুন

ଶୁକ୍ଳ

চলো

অনুসন্ধান

ভিত্তিক

- সংযোগকারী পৃষ্ঠাসমূহ
সম্পর্কিত পরিবর্তন
বিশেষ পৃষ্ঠাসমূহ
ছাপার যোগ্য সংস্করণ
স্বামী সংযোগ
এই নিবন্ধটি উদ্ধৃত
করুন

অন্যান্য ভাষাসমূহ

- Aragonés
العربية
অসমীয়া
Azərbaycan
Беларуская
Български
བོད་སྐད་
Brezhoneg

 প্রবেশ/নতুন অ্যাকাউন্ট

निबन्ध

আলোচনা

प्रस्थापना करण

ইতিহাস

৯ How to read and write in Unicode Bangla

 **Must-read for contributors**

କଳକାତା

Kolkata

স্থানাংক: 22.5697, 88.3697

উৎকলসিঙিয়া, মক্ত বিশ্বকোষ থেকে

কলকাতা (ইংরেজি ভাষায়: Kolkata, পূর্বে **Calcutta** সহচিহ্নিত) ভারতের পূর্বাঞ্চলীয় রাজ্য পশ্চিমবঙ্গের

রাজধানী। হুগলী নদীর পূর্ব তীরে অবস্থিত এই শহরের (পৌর অঞ্চলের) জনসংখ্যা ৫০ লাখের কিছু বেশি।
কলকাতা ও তার পার্শ্ববর্তী এলাকায় কলকাতার মতো একটি শহর আছে।
কাছাকাছি: এই জনসংখ্যার প্রায় ৫০% মানুষ কলকাতার চতুর্থ বর্ডারের মধ্যে বাস করে।
মহানগরীয় অঞ্চল।

১৯১১ **area** তা ব্রিটিশ ভারতের রাজধানী ছিল। একদা ভারতের আধুনিক শিক্ষা, বিজ্ঞান, শিল্প, সংস্কৃতি ও রাষ্ট্রকলকাতা মহানগরে ১৯৫৪ সালের পর থেকে তীব্র রাজনৈতিক সংঘাত ও তার সঙ্গে সঙ্গে আর্থনৈতিক দমকা দমকা ঘটেছে। তবে ২০০০ সালের পর থেকে এই শহর আবার আর্থিক ও বাণিজ্যিক সমৃদ্ধির গণ্ডি অতিক্রম হতে শুরু করেছে।

কলকাতা শহরের **প্রসিদ্ধি** তার বেল্লিখিল আন্দোলনগুলির জন্য। **ভারতের স্বাধীনতা আন্দোলন** এবং প
বামপন্থী **আন্দোলন** এ শহরের ঐতিহ্যবাহী একটি বড় অংশ। এর সঙ্গে সঙ্গে **weight**
ভারতে **famous** **মহানী ও আনন্দ নগরী** (City of Joy) নামে পরিচিত। **নন্দী** নামক একটি
ঠাকুর, **বসু**, মান্নার ভেটেরা, সত্যজিৎ রায়, **part of** **বিবেকানন্দ রাজা**
রামমোহন রায় সি ভি রামান-সহ বহু বিশিষ্টত ব্যক্তিদের বাসভূমি। **কলকাতা** **ঐতিহ্য ও সমৃদ্ধ**
সাংস্কৃতিক প্রেক্ষাপটের কারণে আজও বিশ্বব্যাপী চাওয়া মধ্যবর্তী আসান আশ্রিত।

সৃষ্টিপত্র [আডালে রাখো]

- ১ নামকরণ
- ২ ইতিহাস
- ৩ ভূগোল
- ৩.১ শহরের গঠন
- ৪ অর্থনীতি
- ৫ রাজধানী ও নগর প্রশাসন
- ৬ বাতাস
- ৭ জনপরিসংখ্যান
- ৮ সংস্কৃতি
- ৯ শিক্ষা
- ১০ গণমাধ্যম

weighted sum for PERSON, ws_p
 weighted sum for ORGANIZATION, ws_o
 weighted sum for LOCATION, ws_l

weight = 3

located

population

area

famous

part of

weight = 1

କଳକାଢ଼ା

ଅନ୍ଧାରବଦ୍ଧ : ଭାବତ



জিস্টেডিয়া মোজারিয়া



शुद्धांक: 22.5697, 88.3697

IST (UTC+05:30)

१,५०० वर्ग किमि (५७९ sq mi)

• 2 m (60 ft)

কলকাতা (ওলা

11 8,470,488 (2002)

- **ସମ୍ପଦ**

২.১১৪। (প্রশ্নাংশন-কৃষ্টি: ৫-এই চনা অপবদ্য নহে) ০

Generating an Annotated List of Phrases (contd.)

Steps

- ➊ Generate and annotate the title of each article
 - (i) Use category information to annotate the title
 - (ii) Use a small set of keywords to annotate the title
- ➋ Getting more location names
- ➌ Generating and annotating the tokens in the titles

Generating an Annotated List of Phrases (contd.)

Annotating the tokens in the titles

- Assign **each** token the same NE label as that of its title
- **Ambiguous case:**

Anna University
ORG ORG

Anna Frank
PER PER

- **Solution:** Label “Anna” with its most frequent NE class

Applying the Annotated List to a Text

Example

Smith College is in Massachusetts

Applying the Annotated List to a Text

Example

Smith College is in Massachusetts

Applying the Annotated List to a Text

Example

Smith College is in Massachusetts

ORG

ORG

Applying the Annotated List to a Text

Example

<i>Smith College</i>	<i>is</i>	<i>in</i>	<i>Massachusetts</i>
ORG	ORG	OTHERS OTHERS	LOC

Outline

- 1 Two New Linguistic Features
 - Induced Affixes
 - Semantic Classes from Wikipedia
- 2 POS Tagging with New Features
 - Evaluation
- 3 Pipelined NER with New Features
 - Evaluation
- 4 Joint Model for POS Tagging and NER
 - Evaluation

POS Tagging

Goal

Investigate whether the two new features can improve a baseline supervised POS tagger

Experimental setup

- **Corpus**: 78K word tokens
- **Tagset**: IIIT Hyderabad's POS tagset with 26 tags
- **Learning algorithm**: CRF
- 5-fold cross-validation (CV)

Features

Baseline

- word n-grams, pseudo-affixes

Results

Experiment	Overall	Seen	Unseen
Baseline	89.8	92.9	72.0

Table: 5-fold cross-validation accuracies

Features

Baseline+Induced Affixes

- word n-grams, pseudo-affixes
- induced affix n-grams

Results

Experiment	Overall	Seen	Unseen
Baseline	89.8	92.9	72.0
Baseline+Induced Affixes	90.5	93.3	74.6

Table: 5-fold cross-validation accuracies

Features

Baseline+Induced Affixes+Wiki

- word n-grams, pseudo-affixes
- induced affix n-grams
- induced Wiki n-grams

Results

Experiment	Overall	Seen	Unseen
Baseline	89.8	92.9	72.0
Baseline+Induced Affixes	90.5	93.3	74.6
Baseline+Induced Affixes+Wiki	90.8	93.5	75.5

Table: 5-fold cross-validation accuracies

Outline

- 1 Two New Linguistic Features
 - Induced Affixes
 - Semantic Classes from Wikipedia
- 2 POS Tagging with New Features
 - Evaluation
- 3 **Pipelined NER with New Features**
 - **Evaluation**
- 4 Joint Model for POS Tagging and NER
 - Evaluation

NER

Goal

Investigate whether the two features can improve a **pipelined** NE recognizer that uses induced POS tags as features

Experimental setup

- **Corpus**: 78K word tokens
- **Learning algorithm**: CRF
- IOB convention
- 5-fold cross-validation (CV)

Features

Baseline

- word n-grams, pseudo-affixes, **POS n-grams**

Results

Experiment	R	P	F
Baseline	60.9	74.4	67.0
Person	66.1	74.0	69.9
Organization	29.8	44.9	35.8
Location	52.6	80.4	63.6

Table: 5-fold cross-validation results

Features

Baseline+Induced Affixes

- word n-grams, pseudo-affixes, **POS n-grams**
- induced affix n-grams

Results

Experiment	R	P	F
Baseline	60.9	74.4	67.0
Person	66.1	74.0	69.9
Organization	29.8	44.9	35.8
Location	52.6	80.4	63.6
Baseline+Induced Affixes	60.4	73.3	66.2
Person	65.7	72.6	69.0
Organization	31.7	46.4	37.7
Location	51.4	80.0	62.6

Table: 5-fold cross-validation results

Features

Baseline+Induced Affixes+Wiki

- word n-grams, pseudo-affixes, **POS n-grams**
- induced affix n-grams
- induced Wiki n-grams

Results

Experiment	R	P	F
Baseline	60.9	74.4	67.0
Person	66.1	74.0	69.9
Organization	29.8	44.9	35.8
Location	52.6	80.4	63.6
Baseline+Induced Affixes	60.4	73.3	66.2
Person	65.7	72.6	69.0
Organization	31.7	46.4	37.7
Location	51.4	80.0	62.6
Baseline+Induced Affixes+Wiki	63.2	75.1	68.7
Person	66.4	75.1	70.5
Organization	30.7	43.8	36.1
Location	60.0	79.6	68.5

Table: 5-fold cross-validation results

Results

Experiment	R	P	F
Baseline	60.9	74.4	67.0
Person	66.1	74.0	69.9
Organization	29.8	44.9	35.8
Location	52.6	80.4	63.6
Baseline+Induced Affixes	60.4	73.3	66.2
Person	65.7	72.6	69.0
Organization	31.7	46.4	37.7
Location	51.4	80.0	62.6
Baseline+Induced Affixes+Wiki	63.2	75.1	68.7
Person	66.4	75.1	70.5
Organization	30.7	43.8	36.1
Location	60.0	79.6	68.5

Table: 5-fold cross-validation results

Results

Experiment	R	P	F
Baseline	60.9	74.4	67.0
Person	66.1	74.0	69.9
Organization	29.8	44.9	35.8
Location	52.6	80.4	63.6
Baseline+Induced Affixes	60.4	73.3	66.2
Person	65.7	72.6	69.0
Organization	31.7	46.4	37.7
Location	51.4	80.0	62.6
Baseline+Induced Affixes+Wiki	63.2	75.1	68.7
Person	66.4	75.1	70.5
Organization	30.7	43.8	36.1
Location	60.0	79.6	68.5

Table: 5-fold cross-validation results

Outline

- 1 Two New Linguistic Features
 - Induced Affixes
 - Semantic Classes from Wikipedia
- 2 POS Tagging with New Features
 - Evaluation
- 3 Pipelined NER with New Features
 - Evaluation
- 4 **Joint Model for POS Tagging and NER**
 - **Evaluation**

Joint Model

Goal

Jointly predict a word's POS and NE tag to help avoid error propagation

Model assumption

A Bengali word is part of an NE if and only if it is a proper noun. For our evaluation corpus, this assumption is correct **97.3%** of the time.

Joint Model (contd.)

Model

- Jointly predicts a word's POS and NE tag
- Trained using CRF
 - Features**: POS and NE taggers' features **minus** NE tagger's POS-related features
 - Joint tag**: If a word is not a proper noun, its class is simply its POS tag. Otherwise, its class is its NE tag

Example:

<i>Dhaka</i>	<i>is</i>	<i>the</i>	<i>capital</i>	<i>of</i>	<i>Bangladesh</i>
B-LOC	VBZ	DT	NN	IN	B-LOC

NER Results for Joint Modeling

Experiment	R	P	F
Baseline	54.7	81.7	65.5
Baseline+Induced Affixes	56.7	88.9	69.3
Baseline+Induced Affixes+Wiki	61.7	86.3	71.9

Table: 5-fold cross-validation joint modeling results for NER

Comparison between Joint and Pipelined Model

Experiment	R	P	F
Baseline	54.7	81.7	65.5
Baseline+Induced Affixes	56.7	88.9	69.3
Baseline+Induced Affixes+Wiki	61.7	86.3	71.9

Table: 5-fold cross-validation joint modeling results for NER

Experiment	R	P	F
Baseline	60.9	74.4	67.0
Baseline+Induced Affixes	60.4	73.3	66.2
Baseline+Induced Affixes+Wiki	63.2	75.1	68.7

Table: 5-fold cross-validation results for NER

Summary

- Investigate the pipelined NER architecture and proposed two new features:
 - induced affixes
 - semantic class information from Wikipedia
- Proposed a joint model for learning POS tagging and NER simultaneously