­­МІНІСТЕРСТВО ОСВІТИ ТА НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ “ЛЬВІВСЬКА ПОЛІТЕХНІКА”

­­­



Контрольна робота №1

з курсу «Технології захисту інформації»

для студентів базового напрямку 6.08.04 "Комп’ютерні науки"

(заочна форма навчання)

Варіант 8

Виконав студент гр. КНз-3

Чалий Михайло

­­

Львів 2015

## Індивідуальне завдання

Написати програму для статистичного аналізу тексту. Статистичний аналіз повинен дати такі параметри аналізованого тексту:

- визначити використаний алфавіт;

- частоти повторень одного символу для всіх символів тексту, які представити двома способами – в алфавітному порядку і по спаданню частоти повторення у вигляді гістограм;

- частоти повторень для біграм (двох символів, які зустрічаються в тексті в кількості 10-15), які представити у вигляді гістограм;

- частоти повторень для триграм (трьох символів, які зустрічаються в тексті в кількості 10-15), які представити у вигляді гістограм;

- знайти повторення символів в тексті для 2, 3 і 4символів.

Отримати відкритий текст і виконати статистичний аналіз. Зробити висновки на основі отриманих статистичних характеристик відкритого тексту.

## Статистичний аналіз

Статистичний аналіз відкритого тексту проводиться для того, щоб знайти та виділити частоти повторень одної літери, біграм та триграм. Це дає нам змогу виявити певні властивості та особливості мови. Виконавши статистичний аналіз шифрованого тексту ми можемо виявити певні закономірності, побачити які насправді слова можуть ховатися під шифрованим текстом. При співставленні отриманих даних після аналізу відкритого та шифрованого текстів робляться певні висновки та висуваються гіпотези щодо правильності розшифрування.

## Текст для статистичного аналізу

IT INTRODUCED A RADICALLY NEW METHOD OF DISTRIBUTING CRYPTOGRAPHIC KEYS, WHICH WENT FAR TOWARD SOLVING ONE OF THE FUNDAMENTAL PROBLEMS OF CRYPTOGRAPHY, KEY DISTRIBUTI ON, AND HAS BECOME KNOWN AS DIFFIE-HELLMAN KEY EXCHANGE. THE ARTICLE ALSO STIMULATED THE ALMOST IMMEDIATE PUBLIC DEVELOPMENT OF A NEW CLASS OF ENCIPHERING ALGORITHMS, THE ASYMMETRIC KEY ALGORITHMS. PRIOR TO THAT TIME, ALL USEFUL MODERN ENCRYPTION ALGORITHMS HAD BEEN SYMMETRIC KEY ALGORITHMS, IN WHICH THE SAME CRYPTOGRAPHIC KEY IS USED WITH THE UNDERLYING ALGORITHM BY BOTH THE SENDER AND THE RECIPIENT, WHO MUST BOTH KEEP IT SECRET. ALL OF THE ELECTROMECHANICAL MACHINES USED IN WWII WERE OF THIS LOGICAL CLASS, AS WERE THE CAESAR AND ATBASH CIPHERS AND ESSENTIALLY ALL CIPHER SYSTEMS THROUGHOUT HISTORY. THE 'KEY' FOR A CODE IS, OF COURSE, THE CODEBOOK, WHICH MUST LIKEWISE BE DISTRIBUTED AND KEPT SECRET, AND SO SHARES MOST OF THE SAME PROBLEMS IN PRACTICE. OF NECESSITY, THE KEY IN EVERY SUCH SYSTEM HAD TO BE EXCHANGED BETWEEN THE COMMUNICATING PARTIES IN SOME SECURE WAY PRIOR TO ANY USE OF THE SYSTEM THE TERM USUALLY USED IS 'VIA A SECURE CHANNEL' SUCH AS A TRUSTWORTHY COURIER WITH A BRIEFCASE HANDCUFFED TO A WRIST, OR FACE-TO-FACE CONTACT, OR A LOYAL CARRIER PIGEON. THIS REQUIREMENT IS NEVER TRIVIAL AND VERY RAPIDLY BECOMES UNMANAGEABLE AS THE NUMBER OF PARTICIPANTS INCREASES, OR WHEN SECURE CHANNELS AREN'T AVAILABLE FOR KEY EXCHANGE, OR WHEN, AS IS SENSIBLE CRYPTOGRAPHIC PRACTICE, KEYS ARE FREQUENTLY CHANGED.THE AGING DES WAS OFFICIALLY REPLACED BY THE ADVANCED ENCRYPTION STANDARD AES IN - WHEN NIST ANNOUNCED FIPS. AFTER AN OPEN COMPETITION, NIST SELECTED RIJNDAEL, SUBMITTED BY TWO BELGIAN CRYPTOGRAPHERS, TO BE THE AES. DES, AND MORE SECURE VARIANTS OF IT SUCH AS TRIPLE DES, ARE STILL USED TODAY, HAVING BEEN INCORPORATED INTO MANY NATIONAL AND ORGANIZATIONAL STANDARDS. HOWEVER, ITS X-BIT KEY-SIZE HAS BEEN SHOWN TO BE INSUFFICIENT TO GUARD AGAINST BRUTE FORCE ATTACKS ONE SUCH ATTACK, UNDERTAKEN BY THE CYBER CIVIL-RIGHTS GROUP ELECTRONIC FRONTIER FOUNDATION IN-, SUCCEEDED IN- HOURS. AS A RESULT, USE OF STRAIGHT DES ENCRYPTION IS NOW WITHOUT DOUBT INSECURE FOR USE IN NEW CRYPTOSYSTEM DESIGNS, AND MESSAGES PROTECTED BY OLDER CRYPTOSYSTEMS USING DES, AND INDEED ALL MESSAGES SENT SINCE USING DES, ARE ALSO AT RISK. REGARDLESS OF DES' INHERENT QUALITY, THE DES KEY SIZE WAS THOUGHT TO BE TOO SMALL BY SOME EVEN IN-, PERHAPS MOST PUBLICLY BY WHITFIELD DIFFIE. THERE WAS SUSPICION THAT GOVERNMENT ORGANIZATIONS EVEN THEN HAD SUFFICIENT COMPUTING POWER TO BREAK DES MESSAGES; CLEARLY OTHERS HAVE ACHIEVED THIS CAPABILITY.THE SECOND DEVELOPMENT, IN-, WAS PERHAPS EVEN MORE IMPORTANT, FOR IT FUNDAMENTALLY CHANGED THE WAY CRYPTOSYSTEMS MIGHT WORK. THIS WAS THE PUBLICATION OF THE PAPER NEW DIRECTIONS IN CRYPTOGRAPHY BY WHITFIELD DIFFIE AND MARTIN HELLMAN. IT INTRODUCED A RADICALLY NEW METHOD OF DISTRIBUTING CRYPTOGRAPHIC KEYS, WHICH WENT FAR TOWARD SOLVING ONE OF THE FUNDAMENTAL PROBLEMS OF CRYPTOGRAPHY, KEY DISTRIBUTION, AND HAS BECOME KNOWN AS DIFFIE-HELLMAN KEY EXCHANGE. HE MID-S SAW TWO MAJOR PUBLIC ADVANCES. FIRST WAS THE PUBLICATION OF THE DRAFT DATA ENCRYPTION STANDARD IN THE U.S. FEDERAL REGISTER ON MARCH. THE PROPOSED DES CIPHER WAS SUBMITTED BY A RESEARCH GROUP AT IBM, AT THE INVITATION OF THE NATIONAL BUREAU OF STANDARDS NOW NIST, IN AN EFFORT TO DEVELOP SECURE ELECTRONIC COMMUNICATION FACILITIES FOR BUSINESSES SUCH AS BANKS AND OTHER LARGE FINANCIAL ORGANIZATIONS. AFTER 'ADVICE' AND MODIFICATION BY NSA, ACTING BEHIND THE SCENES, IT WAS ADOPTED AND PUBLISHED AS A FEDERAL INFORMATION PROCESSING STANDARD PUBLICATION IN-. DES WAS THE FIRST PUBLICLY ACCESSIBLE CIPHER TO BE 'BLESSED' BY A NATIONAL AGENCY SUCH AS NSA. THE RELEASE OF ITS SPECIFICATION BY NBS STIMULATED AN EXPLOSION OF PUBLIC AND ACADEMIC INTEREST IN CRYPTOGRAPHY. THE AGING DES WAS OFFICIALLY REPLACED BY THE ADVANCED ENCRYPTION STANDARD AES IN- WHEN NIST ANNOUNCED FIPS-. AFTER AN OPEN COMPETITION, NIST SELECTED RIJNDAEL, SUBMITTED BY TWO BELGIAN CRYPTOGRAPHERS, TO BE THE AES. DES, AND MORE SECURE VARIANTS OF IT, ARE STILL USED TODAY, HAVING BEEN INCORPORATED INTO MANY NATIONAL AND ORGANIZATIONAL STANDARDS. HOWEVER, ITS X-BIT KEY-SIZE HAS BEEN SHOWN TO BE INSUFFICIENT TO GUARD AGAINST BRUTE FORCE ATTACKS ONE SUCH ATTACK, UNDERTAKEN BY THE CYBER CIVIL-RIGHTS GROUP ELECTRONIC FRONTIER FOUNDATION IN-, SUCCEEDED IN- HOURS. AS A RESULT, USE OF STRAIGHT DES ENCRYPTION IS NOW WITHOUT DOUBT INSECURE FOR USE IN NEW CRYPTOSYSTEM DESIGNS, AND MESSAGES PROTECTED BY OLDER CRYPTOSYSTEMS USING DES, AND INDEED ALL MESSAGES SENT SINCE USING DES, ARE ALSO AT RISK. REGARDLESS OF DES' INHERENT

Проведемо аналіз англомовного відкритого тексту для того, щоб дослідити особливості англійської мови. Проведемо цей аналіз за допомогою розробленої програми.

## Алфавіт тексту:

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z \_ . , ; - '

У відкритому тексті було використано всі літери англійського алфавіту (26 літер) та 6 розділових знаків ( \_ . , ; - ' ).

## Аналіз англомовного відкритого тексту

Під час аналізу були отримані наступні графіки (Рис.4.1.1 – Рис.4.1.6).

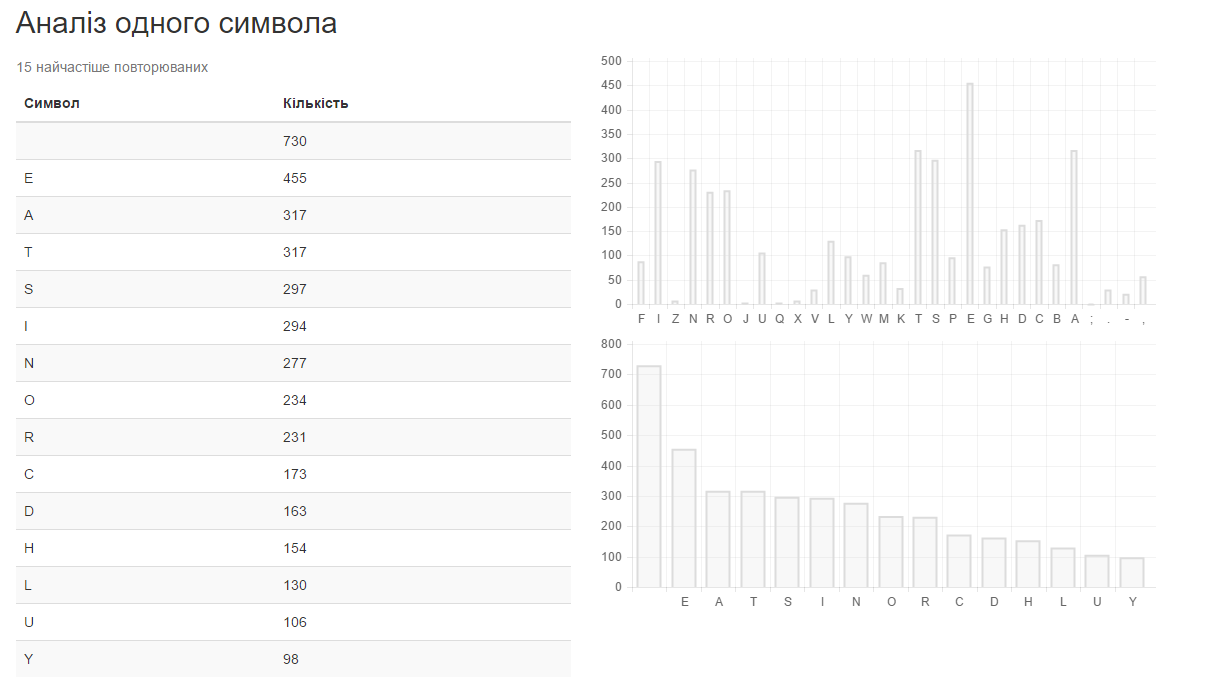


Рис. Аналіз одного символу

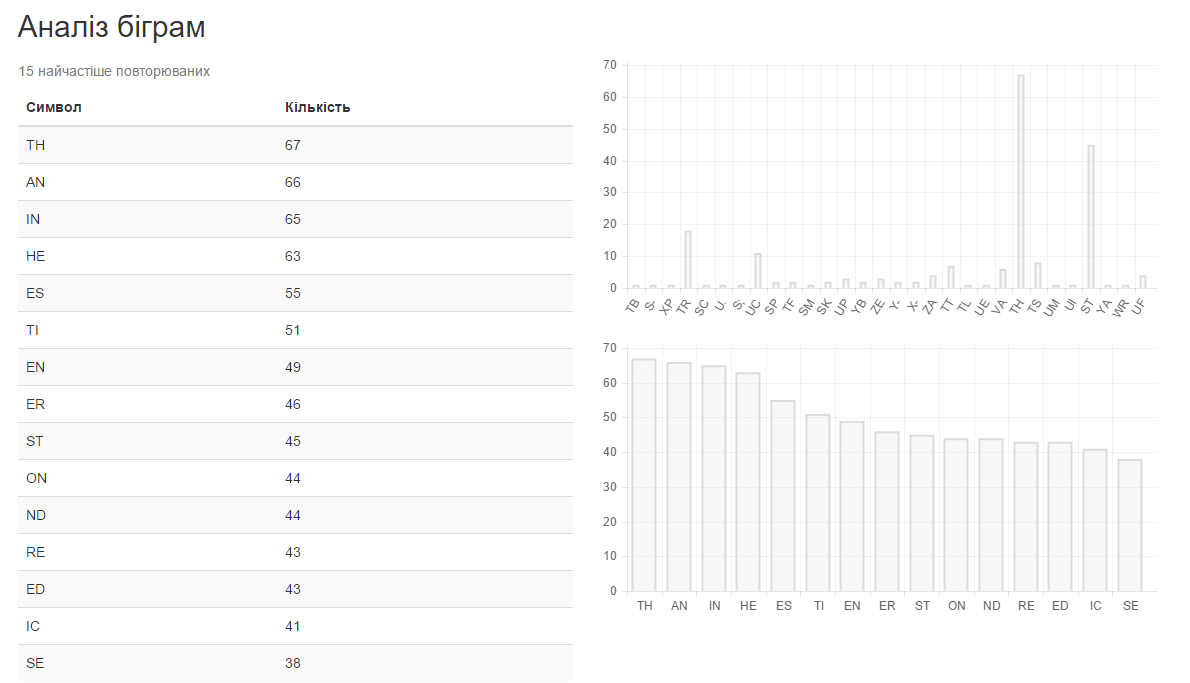


Рис. . Аналіз біграм

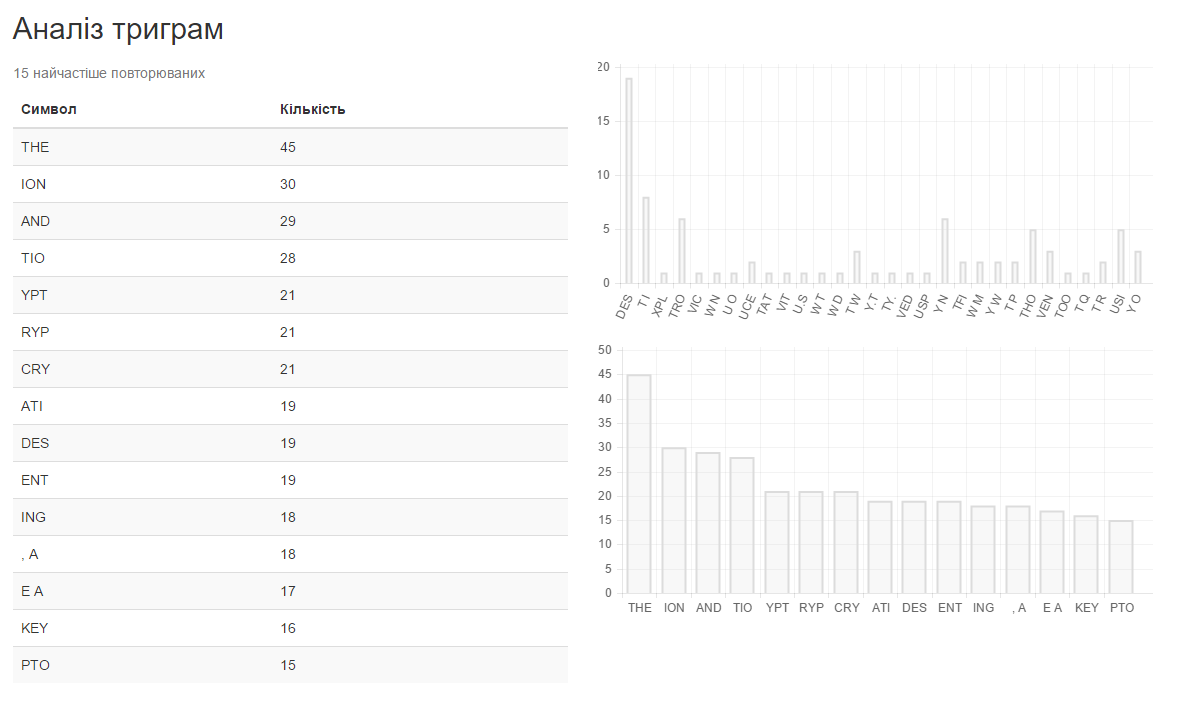


Рис. . Аналіз триграм

З рис на якому зображена гістограма частоти повторень одного символу, бачимо, що використаний наступний алфавіт:

\_ . , ; - ' A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

З рис на якому зображена гістограма частоти повторень одного символу (у порядку спаданню), бачимо, що використаний той самий алфавіт, що і на попередній гістограмі, але посортовану у порядку спадання:

\_ E A T S I N O R C D H L U Y P F M B G W , K V . – ‘ X Z J Q ;

З рис на якому зображена гістограма 15-ти найчастіше повторюваних біграм, бачимо, що в англомовному тексті найчастіше зустрічаються наступні біграми: ,\_ \_A \_I \_S \_T AN D\_ E\_ ES HE IN N\_ S\_ T\_ TH

З рис на якому зображена гістограма 15-ти найчастіше повторюваних біграм у порядку спадання: E\_ \_A S\_ \_T D\_ TH AN IN \_S HE N\_ \_I ,\_ ES T\_

З рис на якому зображена гістограма 15-ти найчастіше повторюваних триграм, бачимо, що в англомовному тексті найчастіше зустрічаються наступні триграми:

\_AN \_DE \_IN \_OF \_TH AND AS\_ ED\_ HE\_ ION ND\_ OF\_ THE TIO YPT

На рис зображена гістограма 15-ти найчастіше повторюваних триграм у порядку спадання:

\_TH THE HE\_ \_IN ED\_ ION \_OF AND \_AN TIO OF\_ AS\_ ND\_ \_DE CRY

З отриманих гістограм можемо визначити особливості англомовного тексту на тематику шифрування. З Рис.4.1.2 бачимо, що найчастіше зустрічається символ пробілу ( \_ , 730 разів повторюється), що зрозуміло, адже пробіл – розділяє слова одне від одного і їх майже така сама кількість як і слів у тексті.

Друга найчастіше повторювана літера – Е (зустрічається 455 раз), яка є частиною найчастіше повторюваних *біграм* Е\_ (означає, що літера Е найчастіше є закінченням слова), НЕ (може бути як окреме слово – займенник, так і її частина,), ES (закінчення слів 3 особи однини в англійській мові) та *триграм* THE (артикль – окреме слово або як частина слова, наприклад слова ), \_ED (закінчення слова в минулому часі), \_DE (початок слова “DES” у даному тексті).

Треті найчастіше повторювані літери – А та Т, які зустрічається 317 разів в тексті. Літера А найчастіше вживається в англійській мові на початку слова ( \_А), як частина окремих слів – AN (артикль), AS («як») та AND (сполучник «і»). Літера Т зустрічається найчастіше на початку ( \_Т, TH) та вкінці слова (Т\_), є частиною артикля THE, частиною суфікса TIO.

Наступні найчастіше повторювані літери – S, яка найчастіше зустрічається в кінці та на початку слова, у біграмах ES, ST,SH, та літера І, яка найчастіше знаходиться у біграмах IN, TI, IT, IO та триграмі TIO.

Літера N найчастіше знаходиться у наступних біграмах: AN, IN, ND та ON, та літера О, яка зустрічається у ON, IO, OF.

Якщо проаналізувати всі шість гістограм (частоту входжень одної літери, біграм та триграм), то можна помітити, що найчастіше повторювані літери E, A, S, T, D та пробіл складають найчастіше повторювані біграми (E\_ \_A S\_ \_T D\_ ) в певних комбінаціях. Такі літери як бачимо найчастіше зустрічаються на початку та в кінці слів.

Наступні біграми зустрічаються досить часто: TH (є частиною триграми (слова) THE або зустрічається в слові), AN (як артикль, частина слова), IN (прийменник або частина слова), HE (закінчення слова або займенник), ES (закінчення слів 3 особи однини в англійській мові), ED (закінчення минулого часу), OF (зазвичай як автономне слово, але також зустрічається як початок слова).

Найчастіше повторювані триграми: TH\_(закінчення слова), THE(артикль, іноді на початку слова), AND (сполучник, іноді зустрічається в словах), ION (закінчення слова).

### Код програми для статистичного аналізу на мові JavaScript

Лістнг analysis-model.js

var EventEmitter = require('events').EventEmitter;

var assign = require('object-assign');

var CHANGE\_EVENT = 'change';

function calculateStats(t){

var onegram = {};

var bigram = {};

var trigram = {};

var fourgram = {};

for(var i = 0; i < t.length; i++){

var char = t.charAt(i);

onegram[char] = (onegram[char] || 0) + 1;

if (i < t.length - 1){

var char2 = char + t.charAt(i+1);

if (char2.trim().length == 2){

bigram[char2] = (bigram[char2] || 0) + 1;

}

if (i < t.length - 2){

var char3 = char2 + t.charAt(i+2);

if (char3.trim().length == 3){

trigram[char3] = (trigram[char3] || 0) + 1;

}

if (i < t.length - 3){

var char4 = char3 + t.charAt(i+3);

if (char4.trim().length == 4){

fourgram[char4] = (fourgram[char4] || 0) + 1;

}

}

}

}

}

var rows = function(hash){

return Object.keys(hash).map(function (key) {

return { name: key, val: hash[key] };

});

};

return {

onegram: rows(onegram),

bigram: rows(bigram),

trigram: rows(trigram),

fourgram: rows(fourgram)

};

}

var \_text = 'YGBPGGFBOQTGBUPQYBHQTBDGVVGTBUMKKPI';

var AnalysisModel = assign({}, EventEmitter.prototype, {

pushTextChanges: function(t){

\_text = t;

this.emitChange();

},

getState: function(){

return assign({

text: \_text

}, calculateStats(\_text));

},

emitChange: function() {

this.emit(CHANGE\_EVENT);

},

addChangeListener: function(callback) {

this.on(CHANGE\_EVENT, callback);

},

removeChangeListener: function(callback) {

this.removeListener(CHANGE\_EVENT, callback);

}

});

if (typeof exports == 'undefined') {

exports = {};

}

exports.calculateStats = calculateStats;

exports.AnalysisModel = AnalysisModel;

Лістінг analysis.js

var React = require('react');

var AnalysisModel = require('../models/analysis-model').AnalysisModel;

var Stats = require('./stats').Stats;

var Analysis = React.createClass({

getInitialState: function() {

return AnalysisModel.getState();

},

reloadState: function(){

this.setState(AnalysisModel.getState());

},

componentDidMount: function() {

AnalysisModel.addChangeListener(this.reloadState);

},

componentWillUnmount: function() {

AnalysisModel.removeChangeListener(this.reloadState);

},

render: function() {

return <form>

<div className='form-group'>

<label htmlFor='inputText'>Текст</label>

<textarea id='inputText'

className='form-control'

rows='10'

value={this.state.text}

onChange={this.\_handleTextChange}></textarea>

</div>

<h2>Аналіз одного символа</h2>

<Stats data={this.state.onegram} />

<h2>Аналіз біграм</h2>

<Stats data={this.state.bigram} />

<h2>Аналіз триграм</h2>

<Stats data={this.state.trigram} />

<h2>Аналіз чотирьох символів</h2>

<Stats data={this.state.fourgram} />

</form>;

},

\_handleTextChange: function(event) {

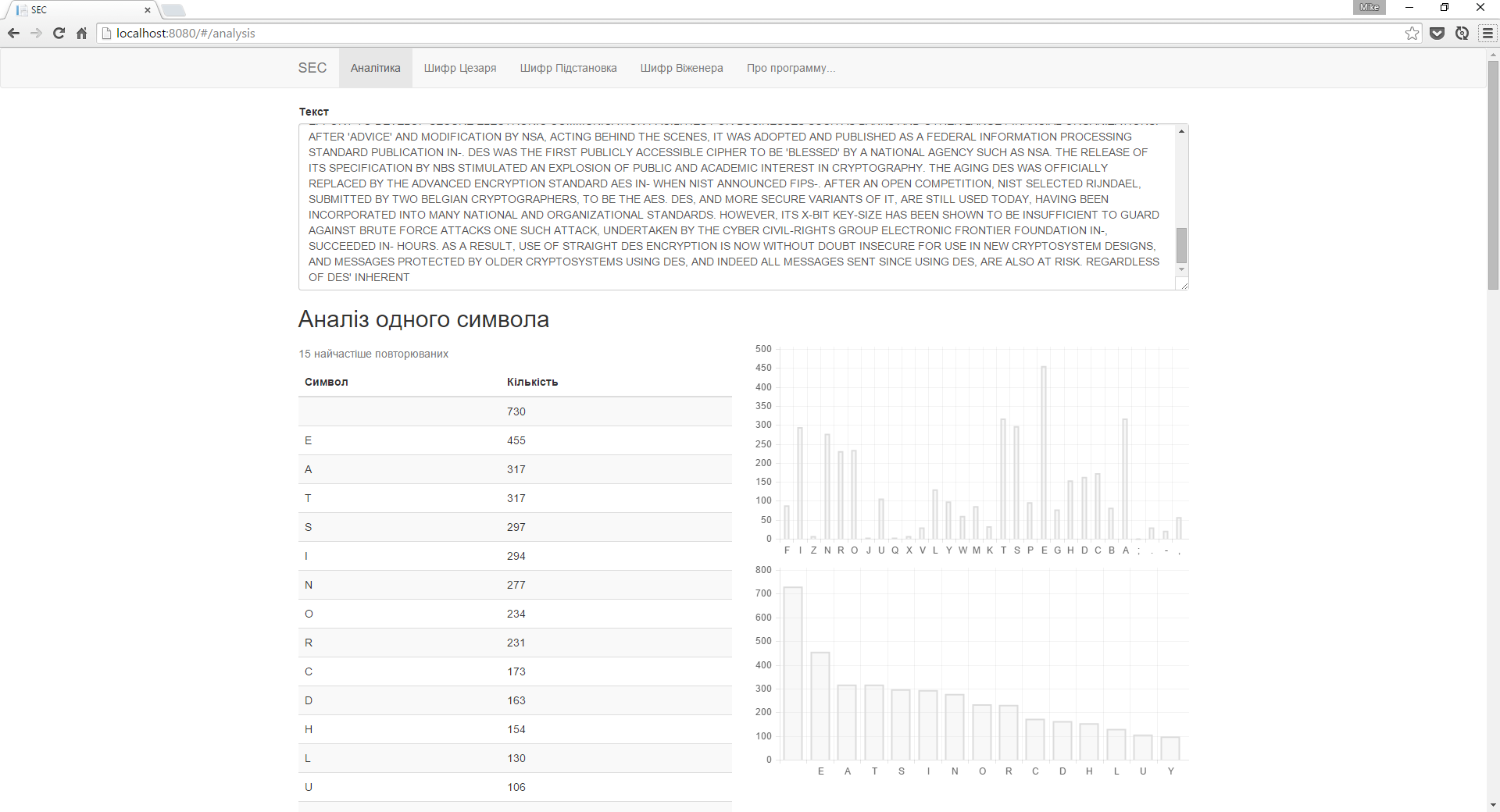
AnalysisModel.pushTextChanges(event.target.value);

}

});

exports.Analysis = Analysis;

## Результат



## Висновок

Виконавши статистичний аналіз ми визначили особливості англомовного тексту на тематику шифрування та дізнались про наступні особливості:

* Найчастіше зустрічається символ пробілу ( \_ , 730 разів), що зрозуміло, адже пробіл – розділяє слова одне від одного.
* Друга найчастіше повторювана літера – Е (455 раз), яка є частиною найчастіше повторюваних біграм Е\_, НЕ, ES та триграм THE, \_ED, \_DE.
* Треті найчастіше повторювані літери – А та Т. Літера А вживається мові на початку слова ( \_А), AN , AS та AND. Літера Т – на початку та вкінці слова, є частиною THE, TIO.
* Біграми зустрічаються досить часто: TH, AN, IN, HE, ES, ED, OF.
* Найчастіше повторювані триграми: TH\_, THE, AND, ION.