НТУУ "КПІ ім Ігоря Сікорського"

Фізико-технічний інститут

КРИПТОГРАФІЯ
КОМП'ЮТЕРНИЙ ПРАКТИКУМ №3

Криптоаналіз афінної біграмної підстановки

Виконали:
студенти групи ФБ-14
Разумний Ілля
Болгов Микола

Перевірила:
Селюх П.В.

Київ 2023

Варіант-1

Мета роботи:

Набуття навичок частотного аналізу на прикладі розкриття моноалфавітної підстановки; опанування прийомами роботи в модулярній арифметиці

Порядок виконання роботи:

Зашифрований текст міститься у файлі to_decrypt_01.txt
Програма razik_bolgov_3.py на мові Python3

На початку виконання лабораторної роботи було написано функції знаходження НСД за розширеним алгоритмом Евкліда, знаходження оберненого за модулем числа та розв'язання конгруенцій

Далі було знайдено найчастіші біграми шифртексту шляхом невеликої модифікації функції із 1 лр. Ось їх словник:

```
{'рн': 0.025368248772504, 'ыч': 0.016775777414075, 'нк': 0.013911620294599, 'цз': 0.013093289689034, 'иа': 0.01227495908347}
```

Після цього у функції *bigram_to_number* переводимо біграми у числове представлення скориставшись формулою

$$(x_{2i-1}, x_{2i}) \leftrightarrow X_i = x_{2i-1}m + x_{2i}.$$

Найскладніша частина лабораторної роботи - написання функції знаходження ключа та дешифрування тексту ключем (перетворення біграм у цифровому вигляді назад у текст)

У нас вишло дуже багато кандидатів на ключ:

```
{(372, 286), (152, 920), (558, 385), (307, 147), (399, 379), (775, 410), (842, 320), (589, 912), (21
41, 540), (589, 651), (450, 47), (837, 227), (744, 664), (806, 496), (796, 711), (241, 821), (327, 7
), (719, 193), (287, 694), (713, 137), (899, 416), (124, 434), (62, 953), (511, 541), (527, 558), (9
, (67, 416), (336, 940), (31, 75), (486, 875), (732, 540), (627, 622), (165, 658), (496, 31), (217,
860), (653, 98), (173, 275), (29, 258), (902, 30), (754, 689), (589, 302), (77, 410), (914, 107), (8
(498, 452), (926, 103), (648, 238), (13, 151), (211, 224), (806, 937), (84, 31), (720, 101), (31, 94
 382), (450, 218), (200, 372), (780, 767), (744, 844), (50, 540), (651, 909), (858, 695), (930, 333)
372, 806), (556, 258), (775, 930), (788, 482), (713, 317), (899, 596), (203, 740), (310, 713), (930,
04, 709), (229, 217), (775, 134), (682, 571), (930, 682), (197, 166), (856, 891), (93, 868), (806, 6
, 165), (341, 444), (62, 602), (248, 881), (707, 897), (465, 726), (713, 837), (421, 647), (651, 224
```

Далі ми вирішили знайти всі можливі тексти та на око оцінити який можливий ключ

Однією з проблем було те, що на початку ми неправильно вказали алфавіт, випадково поміняли місцями “ы” та “ь”, що призвело до того що певні символи були розшифровано неправильно (правильно, але не в тому порядку через індекси алфавіта):

```
Key: (478, 430) with text: ннсгсгауончюбишнисдылостю
Key: (241, 821) with text: аоужужбщзосвфкпожлимьхстя
Key: (389, 885) with text: фржмжмбжирбкыомрдяцпкдстк
Key: (445, 156) with text: еюяфяфкцйюбашйтюнофуэвсто
Key: (864, 519) with text: рлеэеэпзелзхкдулыюйчдастц
Key: (894, 506) with text: бадыдыпгсайилнцазвмцхрстц
Key: (486, 875) with text: цйгцгцрыяйгнщасйькгфсссстю
Key: (929, 651) with text: ухцьцьуежхеябшахнамзмистс
Key: (277, 421) with text: уджзжзеьэдншкхщдмйзэтнстш
Key: (13, 151) with text: многогщуннуюяичностыдосто
Key: (336, 940) with text: кааыаызгыаоиюнаапвкццрств
Key: (666, 797) with text: жшоеоеччлшьлсяшеняьщюсти
Key: (483, 908) with text: сунюнюянрузгюшжуцпыжутноб
```

Правильно було поставити “ь” та “ы”

Після цього нам вивело правильний фрагмент, повний фрагмент далі у протоколі

Для автоматичної перевірки на визначення змістовного тексту ми обрали ***критерій заборонених l-грам***

Текст є змістовним тоді і тільки тоді, коли у ньому відсутні біграми, яких немає у природній мові

Для підрахунку заборонених біграм скористуємось програмою з 1 лр, трохи модифікувавши код:

```
117     df = pd.DataFrame(columns=alphabet, index=alphabet)
118
119     for bigram, freq in bigram_freq_dict.items():
120         if len(bigram) == 2:
121             df.loc[bigram[0], bigram[1]] = freq
122
123     not_allowed_bigrams = []
124     df = df.isna()
125     for i in df:
126         for j in df:
127             print(type(df.loc[i, j]))
128             if df.loc[i, j]:
129                 not_allowed_bigrams.append(i+j)
130
131     print(not_allowed_bigrams)
132
133     df.fillna( value: 0, inplace=True)
134     exit(0)
```

У виводі маємо список біграм із значенням NaN, тобто тих біграм, яких немає у природній російській мові (тобто ми перевели кожну клітинку під False якщо є значення та True якщо немає значення, що нам і треба)

Отримуємо вивід:

not_allowed_bigrams = ['аы', 'аь', 'бй', 'вй', 'гй', 'гф', 'гх', 'гц', 'гш', 'гщ', 'гы', 'гь', 'гю', 'дй', 'дф', 'дщ', 'еы', 'еь', 'жй', 'жм', 'жх', 'жш', 'жщ', 'жы', 'жю', 'зй', 'зщ', 'иь', 'йй', 'йы', 'йь', 'кй', 'кщ', 'кы', 'кь', 'мй', 'нй', 'оы', 'оь', 'пй', 'пц', 'пщ', 'пю', 'рэ', 'сй', 'сщ', 'тй', 'уы', 'уь', 'фг', 'фж', 'фз', 'фй', 'фк', 'фх', 'фц', 'фч', 'фш', 'фщ', 'фэ', 'фю', 'фя', 'хй', 'хщ', 'хы', 'хю', 'цб', 'цж', 'цй', 'цц', 'цш', 'цщ', 'ць', 'цю', 'ця', 'чй', 'чх', 'чц', 'чы', 'чю', 'шг', 'шж', 'шз', 'шй', 'шф', 'шщ', 'шы', 'шэ', 'шя', 'щд', 'щж', 'щз', 'щй', 'щл', 'щм', 'щп', 'щх', 'щц', 'щч', 'щш', 'щы', 'щэ', 'щю', 'щя', 'ыы', 'ыь', 'ьй', 'ьы', 'ьь', 'эа', 'эб', 'эв', 'эг', 'эе', 'эж', 'эз', 'эи', 'эо', 'эу', 'эц', 'эч', 'эщ', 'эы', 'эь', 'ээ', 'эю', 'эя', 'юы', 'юь', 'яы', 'яь']

При запуску тесту на змістовність стикнулись із проблемою, що жоден з текстів не проходить перевірку. Виявилось, що навіть такої класичної книги, як “Дванадцять стільців” недостатньо для вивчення статистичних властивостей мови!
Нам заважала біграма “ЭГ”, яка є у природній російській мові, проте у тексті з 1 лаби вона була відсутньою. Це свідчить про те, що для дослідження властивостей мови відносно невеликих текстів недостатньо

Також видалимо деякі інші біграми, які можуть трапитись у російській мові. Після запуску текст із ключем 13, 151 проходить перевірку. Він пройшов перевірку на заборонені біграми та є змістовним. Це уривок із твору З. Фрейда “Достоєвський та батьковбивство”:

```
Key: (13, 151) with text: многограннуюличностьдостоевскогоможнорассматриватьсчетырехсторонкакписа
```

Повний текст:

Key: (13, 151) with text:
многограннуюличностьдостоевскогоможнорассматриватьсчетырехсторонкакписателякакневротикакакмыслителяэтикаикакгрешникакакжеразобратьсявэтойневольносмущающейнассложностинаименееспоренонкакписательместоеговодномрядусшекспиромбратьякарамазовывеличайшийроманизвсехкогдалибонаписанныхалегендаовеликоминквизитореодноизвысочайшихдостиженийимировойлитературыпереоценитькотороеневозможнокожалениюпередпроблемойписательскоготворчествапсихоанализдолженсложитьоружиедостоевскийскореевсегоуязвимкакморалистпредставляяегочеловекомвысоконравственнымнатомоснованииичтотолькототдостигаетвысшегонравственногосовершенствактопрошелчерезглубочайшиебездныгреховностимыигнорируемодносоображениеведьнравственнымявляетсячеловекреагирующийуженавнутреннеиспытываемоеискушениеприэтомеунеподдаваяськтожепопеременноготорешиттораскаиваясьставитсебевысокиенравственныецелитоголегкоупрекнутьвтомчтоонслишкомудобнодлясебястроитсвоюжизньоннеисполняетосновногопринципанравственностинеобходимостиотреченияавтовремякакнравств

енныйобразжизнивпрактическихинтересахвсегочеловечестваэтимоннапоминаетварваро
вэпохипереселениянародовварваровубивавшихизатемкаявшихсявэтомтакчтопокаяниест
ановилосьтехническимпримеромрасчищавшимпутькновымубийствамтакжепоступалив
ангрозныйэтасделкассовестьюхарактернаярусскаячертадостаточнобесславениконечный
итогнравственнойборьбыдостоевскогопослеисступленнойборьбывоимяпримиренияпри
тязанийпервичныхпозывовиндивидастребованиямичеловеческогообществаонвынужден
норегрессируеткподчинениюмирскомуидуховномуавторитетукпоклонениюцарюихрист
ианскомубогукрусскомумелкодушномунационализмукчемуменеезначительныеумыпри
шлисгораздоменьшимиусилиямичемонвэтомслабоеместобольшойличностидостоевский
упустилвозможностьстатьучителемиосводителемчеловечестваиприсоединилсяктюре
мщикамкультурабудущегонемногимбудетемуобязанавэтомповсейвероятностипроявился
егоневрозиззакоторогоонибылосужденнатакуюнеудачупомощипостиженияисилелюбви
клюдямемубылоткрытдругойапостольскийпутьслужениянампредставляетсяотталкиваю
щимрассматриваниедостоевскоговкачествегрешникаилипреступниканоэтоотталкивани
енедолжноосновыватьсянаобывательскойоценкепреступникавыявитьподлиннуюмотива
циюпреступлениянедолгодляпреступникасущественныдвечертыбезграничноесебялюби
еисильнаядеструктивнаясклонностьобщимдляобеихчертипредпосылкойдляихпроявлен
ийявляетсябезлюбовностьнехваткаэмоциональнооценочногоотношениякчеловекутутср
азувспоминаешьпротивоположноеэтомуудостоевскогоегобольшуюпотребностьвлюбвии
егоогромнуюспособностьлюбитьпроявившуюсявегосверхдобротеипозволявшуюемулю
битьипомогатьтамгдеонимелбыправоненавидетьимститьнапримерпоотношениюкегопе
рвойженеиеелюбовникунотогдавозникаетвопросоткудаприходитсоблазнпричислениядо
стоевскогокпреступникамответиззавыбораегосюжетовэтопреимущественнонасильники
убийцыэгоцентрическиехарактерычтосвидетельствуетосуществованиитакихсклонносте
йвеговнутреннеммиреатакжеиззанекоторыхфактовегожизнистрастиегоказартнымиграм
можетбытьсексуальногорастлениянезрелойдевочкиисповедьэтопротиворечиеразрешает
сяследующимобразомсильнаядеструктивнаяустремленностьдостоевскогокоторыймогла
бысделатьегопреступникомбылавегожизнинаправленаглавнымобразомнасамогосебявов
нутрьвместотогочтобыизнутриитакимобразомвыразилассьвмазохизмеичувствевинывсет
акивеголичностинемалоисадистическихчертвыявляющихсявегораздражительностимуч
ительственетерпимостидажепоотношениюклюбимымлюдяматакжевегоманереобращен
иясчитателемитаквмелочахонсадиствовневажномсадистпоотношениюксамомусебесле
довательномазохистиэтомягчайшийдобродушнейшийвсегдаготовыйпомочьчеловеквсло
жнойличностидостоевскогомывыделилитрифактораодинколичественныйидвакачествен
ныхегочрезвычайноповышеннуюаффективностьегоустремленностькперверзиикотораяд
олжнабылапривестиегоксадомазохизмуилисделатьпреступникомиегонеподдающееесян
ализутворческоедарованиетакоесочетаниевполнемоглобысуществоватьибезневрозаведь
бываютжестопроцентныемазохистыбезналичияневрозовпосоотношениюсилпритязании
первичныхпозывовипротивоборствующихимторможенийприсоединяясюдавозможност
исублимированиядостоевскоговсеещеможнобылобыотнестикразрядуимпульсивныххар
актеровноположениевещейзатемняетсяналичиемневрозанеобязательногокакбылосказан
оприданныхобстоятельствахновсежевозникающеготемскореечемнасыщенноосложнен
иеподлежащеесостороныческогояпреодолениюневрозэтотолькознактогочтоятако
йсинтезнеудалсячтооноприэтойпопыткепоплатилосьсвоимединствомвчемжевстрогомсм

ыслепроявляетсяневроздостоевскийназывалсебясамидругиетакжесчиталиегоэпилептик омнатомоснованиичтоонбылподвержентяжелымприпадкамсопровождавшимисяпотерей сознаниясудорогамиипоследующимупадочнымнастроениемвесьмавероятночтоэтатакна зываемаяэпилепсиябылалишьсимптомомегоневрозакоторыйвтакомслучаеследуетопред елитькакистероэпилепсиютоестькактяжелуюистериюутверждатьэтосполнойуверенност ьюнельзяподвумпричинамвопервыхпотомучтодатыанамнезическихприпадковтакназыва емойэпилепсиидостоевскогонедостаточныиненадежныавовторыхпотомучтопониманиес вязанныхсэпилептоиднымиприпадкамиболезненныхсостоянийостаетсянеясныма

Висновки:

В ході виконання лабораторної роботи було виконано успішну атаку на шифр афінної біграмної підстановки. Вона була виконана шляхом частотного аналізу та використання формул афінної підстановки:

$$Y_i = (aX_i + b) \bmod m^2, \qquad X_i = a^{-1}(Y_i - b) \bmod m^2$$

Існують різні критерії перевірки тексту на змістовність. Особисто нами було обрано критерій *критерій заборонених l-грам.* Він грунтується на тому, що перевірку проходить тільки той текст, у якому відсутні біграми, яких не існує у природній мові. Наприклад, ЖЫ, ШЫ, ЯЬ, АЬ, тощо

Окрім цього, до можливих критеріїв перевірки можна віднести наступні тести:

- перевірка частот частих літер / біграм (у природньому змістованому тексті найчастіше траплятимуться літери о, а, е, и, н та біграми «ст», «но», «то», «на», «ен»
- перевірка частот рідкісних літер (аналогічно до попереднього, тільки навпаки)
- перевірки частот довільних l-грам (наприклад, триграми “что” у російській мові або “the” у англійській

Кошенятко після ЛР 3 Crypto