

Codetheorie: Ontcijfer opdracht

Robin Verachtert - s0121405

26 march 2015

Contents

1	Playfair	4
1.1	De opdracht	4
1.2	Poging 1: frequentie analyse	4
1.3	Poging 2: slimme frequentie analyse	4
1.4	Poging 3: Hill Climb Algoritme	5
1.5	Poging 4: Churn Algoritme	5
1.6	De encrypted text	6
2	ADFGVX	8
2.1	De opdracht	8
2.2	Morse decoderen	8
2.3	De kolomtranspositie	8
2.4	Vinden van de taal.	9
2.5	Vinden van de bron tekst	9
3	Enigma	10
3.1	Opdracht	10
3.2	Een beetje extra onderzoek	11
3.3	Maken van de Enigma machine	11
3.4	Crib graph en begin positie van de rotors	12
3.5	Bepalen van het plugboard en ontcijferen van de tekst	12

De eerste ciphertekst was versleuteld met een Vigenère cipher gevolgd door een enkele kolomtranspositie. Bij het ontcijferen moesten we dus eerst de kolomtranspositie ongedaan maken om een ciphertekst te bekomen die enkel met Vigenère versleuteld was. We zochten een manier om snel te checken of een ciphertekst met Vigenère versleuteld was, zodat we de kolomtranspositie konden bruteforcen in een haalbare tijd. Hiervoor vonden we de "index of coincidence" (IC). Deze wordt als volgt berekend: $IC = \frac{\sum_{i=1}^c n_i(n_i-1)}{N(N-1)/c}$ met n_i de frequenties van de c verschillende letters in een tekst lengte N . In essentie is dit een frequentietabel samengevat in 1 waarde. Hoe groter de IC, hoe groter de kans dat twee willekeurig gekozen letters in een plaintext dezelfde zijn, met $IC = 1$ voor een random tekst waarbij elke letter evenveel voorkomt. Voor een Engelse tekst ligt de IC rond de 1.73. Merk op dat bij deze waarde geen rekening gehouden wordt met welke frequentie bij welke letter hoort. Dit betekent dat een plaintext na een monoalfabetische substitutie zijn IC behoudt. Bij een Vigenère ciphertekst met sleutellengte n wordt op elke letter dezelfde monoalfabetische substitutie toegepast als op elke letter er op kn posities vandaan (k geheel). Als we een sleutellengte gokken, kunnen we de tekst verdelen in verzamelingen letters die met dezelfde substitutie versleuteld zouden zijn. Indien de gegokte lengte correct is, moet (indien de ciphertekst lang genoeg was) elke verzameling een redelijk hoge IC hebben, en moeten deze waarden redelijk dicht bij elkaar liggen. Indien de gegokte lengte fout is, zijn de letters in een verzameling niet met dezelfde monoalfabetische substitutie versleuteld (tenzij de gegokte lengte een veelvoud van de correcte is) en lijkt dit eerder een willekeurige verzameling letters, wat een IC dicht bij 1 zal opleveren. Voor onze ciphertekst konden we dus alle mogelijke kolomtransposities ongedaan maken tot een bepaald aantal kolommen om dan te controleren of een mogelijke Vigenère sleutellengte een hoge IC opleverde. Voor een ciphertekst waarop een enkele kolomtranspositie toegepast is met $k \in [1, n]$ kolommen, zijn $\sum_{k=1}^n k!$ transposities mogelijk. Dit wordt al snel onmogelijk om in een haalbare tijd uit te rekenen, maar gelukkig is het aantal kolommen doorgaans klein.

Het ongedaan maken van de transposities gebeurt voor elke mogelijke permutatie van k kolommen als volgt: de ciphertekst lengte l bestaat uit opeenvolgend de letters uit kolommen 1 t.e.m. k . In elke kolom komen minstens $\lfloor \frac{l}{k} \rfloor$ voor. In de eerste $l \% k$ kolommen (in de volgorde bepaald door het codewoord voor de kolomtranspositie = gepermuteerde kolommen) komt nog 1 extra letter voor, zodat alle kolommen samen uiteindelijk l letters bevatten. We kunnen de kolommen makkelijk opvullen door gewoon de ciphertekst te overlopen en 1 voor 1 de kolommen op te vullen. Daarna overlopen we de kolommen round-robin in gepermuteerde volgorde en halen we telkens de eerstvolgende letter uit de kolom. In deze volgorde vormen de letters de ciphertekst voor de kolomtranspositie. Op elk van deze cipherteksten

voeren we dan, voor elke mogelijke sleutellengte l tot een bepaalde waarde, de volgende Vigenèretest uit: we verdelen de ciphertekst in l groepen door de letters 1 voor 1 round-robin over de groepen te verdelen. Elk van deze groepen zou dan door dezelfde monoalfabetische substitutie versleuteld zijn. We berekenen de IC van elke groep en dan het gemiddelde ervan. Indien dit gemiddelde boven een vooraf ingestelde waarde komt, is dit waarschijnlijk de correctie sleutellengte (of een veelvoud ervan).

Bij het toepassen van dit principe bij de gegeven ciphertekst, moesten we eerst wat spelen met de variabelen (maximale keylengtes voor Vigenère en kolomtranspositie en minimale IC), maar uiteindelijk vonden we (met een runtime van slechts een viertal seconden) zes kolomtransposities van zes kolommen waarbij Vigenère met sleutellengte 8 een IC van ongeveer 2.15 opleverde. Na vergelijken met wat andere zelf berekende IC's van Nederlandstalige teksten, waren we hier al redelijk zeker dat deze plaintekst Nederlands was. Daarna berekenden we bij elk van de zes mogelijke transposities voor elk van de 8 groepen letters de meest voorkomende letter. Aannemend dat dit de E was (in het Nederlands erg waarschijnlijk) konden we meteen de monoalfabetische substitutie ongedaan maken op elke groep. Door de volgorde van de letters te reconstrueren, bekwamen we bij een van de zes transposities de volgende plaintekst, een fragment uit "Erik of het Klein Insectenboek" door Godfried Bomans:

DEKLEINEERIKLAGJUISTOPHETOGENBLIKDATDITBOEKJEB
EGINTINHETOUDEBEDVANGROOTMOEDERPINKSTERBLOMME
TDETROONHEMELENDEZIJDENKWASTENENKEEKOVERDERAN
DVANHETBLANKELAKENDESCHEMERIGEKAMERINHETWASHE
TUURWAAROPDEKLEINEMENSENNAARBEDGAANHETUURWAA
RDEGROTEMENSENNIETVANWETENALLEVERTROUWDEDINGE
NVANDEMUURVERVAGENZOETJESAANINHETGROEIENDEDUIS
TERENDEWERELDWORDTSTILZOSTILDATZIJZELFSNIETMEERA
DEMTBUITENSTAPNOGIEMANDVOORBIJSTAPSTAPZOKLINKT
HETENINDEVERTEROEPTJEONGETJEHOOGENFIJNNAAREEN
ANDERJONGETJEZIJNSTEMKLINKTINDEAVONDENJEDENKTD
ARISTOCHEENJONGETJEOPDEWERELDDATNOGNIETINBEDLIG
TERIKLAGSTILTEKIJKENNAARHETRAAMINDEVERTEENNAARD
ESCHEMERENDEPORTRETTEENVANDEMUURHETISNETDACHTHI
JOFERIETSGEBEURENGAATENMISSCHIENGAATEROOKWELIET
SGEBEURENENHIJBESLOOTOMNUEENSNIETGELIJKOPANDERE
AVONDENINSLAAPTEVALLENMAARGOEDOPTETELETENOFERMI
SSCHIENIETSGEBEURENGINGNUWASDAAREENGOEDMIDDELVO
ORWANTONDERZIJNHOOFDKUSSENLAGREENBOEKJESOLMSBEK
NOPTENATUURLIJKEHISTORIEGEHETENENERIKMOESTDAARV
OORMORGENALLEINSECTENUITKENNENHIJHADERDEZEHELEW
OENSDAGMIDDAGUITZITTENLERENENWASTOTAANDEMEIKEV

ERSGEKOMENMORGENOCHTENDONDERHETSPEELKWARTIERZ
OUHIJDEMEIKEVERSERBIJNEMENLAATEENSKIJKENMOMPELD
EERIKHOEVEELPOTENHEEFTEENWESPOOKALWEERZESDEOGE
NZIJNAPARTVERSTELBAARENSTAANVOORINDEKOPMOOIZIJLE
VENNIETINKORVENGELIJKDEBIJENMAARJAWAARLEVENZIJDA
NZIJZULLENAPARTLEVENDENKIKNUDATDOETEROOKNIETTOE
ZIJBEHORENTOTDEFAMILIEDERVLIESVLEUGELIGENENHEBBE
NGEKNIKTESPRIETENENHOESTAATHETMETDEVLINDERS

1 Playfair

1.1 De opdracht

De tweede ciphertekst was versleuteld met Playfair. We moesten dus de key zien te vinden waarmee een matrix was opgesteld om digrammen te versleutelen. We hadden op dit punt al een Nederlandse en een Franse tekst ontcijferd, dus we gingen ervanuit dat dit Engels was. Het oplossen van Playfair bleek moeilijker als ADFGVX en Vigenere, na 4 gefaalde pogingen is het ons uiteindelijk gelukt.

1.2 Posing 1: frequentie analyse

Onze eerste poging bestond uit een frequentieanalyse van de digrammen. We vergeleken die waarden met gekende waarden voor digrammen in het Engels die we zelf hadden berekend aan de hand van enkele Engelse plain-teksten. Na veel puzzelwerk raakten we echter niet erg ver. Er zijn 600 digrammen in Playfair (hoewel ze niet allemaal in de ciphertekst voorkwamen) en door dicht bij elkaar gelegen frequenties was het bijna onmogelijk ze juist te kiezen. We bekeken ook de frequenties van opeenvolgende digrammen samen en we hielden rekening met frequente digrammen waarvan ook het omgekeerde frequent was.

1.3 Posing 2: slimme frequentie analyse

Aangezien we door gewoon naar de frequenties te kijken toch 2 digrammen vrij zeker wisten ("OS" \leftrightarrow "th", "OB" \leftrightarrow "he"), probeerden we om gebruik te maken van de structuur van het Playfair vierkant¹. Dit werkte echter ook niet, omdat we geen van de andere digrammen echt zeker wisten, en diegenen die we dachten juist te gokken bleken niet in het vierkant te passen.

Dus probeerden we de ciphertekst te bruteforcen met keys waarbij deze reeds gevonden digrammen klopten, maar vonden ook zo het antwoord niet.

¹<http://www.umich.edu/~umich/fm-34-40-2/ch7.pdf>

1.4 Posing 3: Hill Climb Algoritme

Daarna gooiden we het over een andere boeg en gingen we op zoek naar een algoritme om de ciphertekst volledig geautomatiseerd te kraken. Eerst maakten we een implementatie van een hill climb algoritme. We starten met een bepaalde startset van mogelijke keys (in ons geval 100 random gekozen keys), de besten op basis van Index of Coincidence²³ hieruit te kiezen en deze te combineren, en deze combinaties dan terug als nieuwe startset te beschouwen. Op een gegeven moment zitten we in een maximum (lokaal) en dit wordt dan teruggegeven. Doordat we voor de combinatie-stap kozen voor een groot aantal keer 2 letters in een key te wisselen en er dus eigenlijk geen combinatie plaatsvond duurde dit algoritme extreem lang om iets te vinden of belande het direct in een lokaal en zeer slecht maximum.

1.5 Posing 4: Churn Algoritme

Daarna stootten we op het zogenaamde Churn algoritme^{4 5}. Dit algoritme is gelijkaardig aan simulated annealing, alleen heel wat simpeler om te implementeren.

Elke plaintekst krijgt een score toegewezen als volgt: voor elke mogelijke digram in het Engels werd de frequentie geanalyseerd. Hiervan werden telkens de log genomen om de invloed van erg grote waarden te beperken, en werden deze waarden herschaald naar 0-9. Nu wordt voor elk digram (ook twee opeenvolgende letters die volgens Playfair in verschillende digrams zitten!) in de plaintekst de frequentie van 0-9 bij de score van de plaintekst geteld. Hoe hoger de score, hoe waarschijnlijker dat de tekst Engels is dus. Merk hierbij op dat de plainteksten van de gegeven ciphertekst komen, dus erg hoog scorende maar niet-Engelse plainteksten als "eeeeeeeeee" zullen niet voorkomen. In het algoritme starten we met een zogenaamde parent key (bv. gewoon het alfabet) waarmee we de ciphertekst ontcijferen en een score toekennen. Daarna voeren we een kleine verandering door aan de parent key en noemen we het resultaat de child key. Deze verandering kan een horizontale of verticale spiegeling, een combinatie van de twee, een verwisseling van twee willekeurige rijen of kolommen of een verwisseling van twee letters zijn. Hierbij komt de verwisseling van twee letters veel meer voor dan de andere opties, dit omdat er veel meer mogelijke verwisselingen van twee letters zijn. De plaintekst voor de child key wordt ook geëvalueerd. Indien de child key beter scoorde, vervangt deze de parent key. Indien de parent key beter scoorde, wordt een willekeurig getal uit een array van 100

²<http://practicalcryptography.com/cryptanalysis/text-characterisation/index-coincidence/>

³http://en.wikipedia.org/wiki/Index_of_coincidence

⁴<http://www.cryptoden.com/index.php/algorithms/churn-algorithm/20-churn-algorithm>

⁵<http://s13.zetaboards.com/Crypto/topic/6781204/1/>

getallen gekozen. Indien het verschil tussen parent en child key scores minder was dan dit gekozen getal, vervangt de child key toch de parent key. Hierdoor kan het algoritme uit lokale maxima raken. Het algoritme blijft oneindig lopen en print de uitkomst van een iteratie enkel indien een nieuwe topscore bereikt is. Merk op dat de 100 getallen in de genoemde array zodanig gekozen zijn dat de kans dat child parent vervangt, gelijkaardig is aan die bij simulated annealing. Bij sommige runs van het algoritme vonden we al na een tweeduizendtal iteraties een tekst die heel erg op Engels leek. Het antwoord was niet helemaal correct gezien bij de logaritmen van de frequenties van digrammen geen rekening gehouden werd met de meer voorkomende X bij Playfair. Het was wel dicht genoeg bij Engels dat we de laatste aanpassingen handmatig konden doorvoeren. We vonden als key "A brief history of time", met als plaintekst het begin van "A Brief History of Time: From the Big Bang to Black Holes" door Stephen Hawking. Het bijbehorende vierkant kan u vinden in Figure 1 ⁶. Merk op dat we de twee lijsten met waarden op <http://www.cryptoden.com/> vonden, maar de code verder helemaal zelf geschreven is en enkel gebaseerd is op de beschrijving van het algoritme.

Volgens de beschrijving van het algoritme waren de 100 waarden in de array afgesteld op een ciphertekst van 100-120 karakters en moest deze nog geschaald worden voor langere teksten. Wij kregen echter betere resultaten zonder de waarden te schalen.

Het algoritme in playfair.py is een licht aangepaste versie van wat we eerst gebruikten. Het geeft vaker snel een antwoord door rollbacks uit te voeren als een tijdje geen vooruitgang geboekt wordt, of zelfs helemaal opnieuw te beginnen. Ook is het scoren nu aangepast aan de X'en in Playfair waardoor exact de correcte plaintekst wordt gevonden. De gevonden key matrix is niet altijd de matrix gegenereerd met "abriefhistoryoftime", maar wel altijd een correcte matrix. Gezien niet alle digrammen voorkomen, zijn meerdere correcte matrices mogelijk.

In de eerste twee lijntjes van het Churn algoritme wordt een vaste seed ingesteld. Deze levert al na 2406 iteraties het antwoord op. Om met een andere seed te proberen, moeten de eerste twee regels code verwijderd worden. Hierdoor kan het algoritme wel langer duren. De variërende runtime is een gevolg van de niet-deterministische aard van het algoritme. Gezien de code niet concurrent is, is het aangewezen om het programma meerdere keren tegelijk te draaien om sneller tot een resultaat te komen.

1.6 De encrypted text

AWELXLKNOWNSCIENTISTSOMESAYITWASBERTRANDRUSXSEL
XLONCEGAVEAPUBLICLECTUREONASTRONOMYHEDESCRIBED

⁶http://www.fisica.net/relatividade/stephen_hawking_a_brief_history_of_time.pdf#page=3

A	B	R	I	E
F	H	S	T	O
Y	M	D	G	J
K	L	N	P	Q
U	V	W	X	Z

Figure 1: The playfair square with key A brief history of time

HOWTHEXEARHTORBITSAROUNDTHESUNANDHOWTHESUNINT
URNORBITSAROUNDTHECENTEROFAVASTCOLLECTIONOFST
ARSCALXLED OURGALAXYATXTHEXENDOFTHELECTUREALITX
TLEOLDLADYATXTHEBACKOFTHEROXOMGOTUPANDSAIDWH
ATYOUHAVETOLDUSISRUBXBISHTHEWORLDISREALXLYAFLAT
PLATESUPXPORTEDONTHEBACKOFAGIANTXTORTOISETHESCI
ENTISTGAVEASUPERIORSMILEBEFOREREPLYINGWHATISTHET
ORTOISESTANDINGONYOUREVERYCLEVERYOUNGMANVERYCL
EVERSAIDTHEOLDLADYBUTITSTURTLESALXLTHEWAYDOWNM
OSTPEOPLEWOULDFINDTHEPICTUREOFOURUNIVERSEASANIN
FINITETOWEROFTORTOISESRATHERXRIDICULOUSBUTWHYDO
WETHINKWEKNOWBETXTERWHATDOWEKNOWABOUTXTHEU
NIVERSEANDHOWDOWEKNOWITWHERE DIDTHEUNIVERSECOM
EFROMANDWHEREISITGOINGDIDTHEUNIVERSEHAVEABEGINX
NINGANDIFSOWHATHAPXPENEDBEFORETHENWHATISTHENAT
UREOFTIMEWILXLITEVERCOMETOANENDCANWEGOBACKINTI
MERECENTBREAKTHROUGHSINPHYSICSMADEPOSXSIBLEINPA
RTBYFANTASTICNEWTECHNOLOGIESXSUGXGESTANSWERSTO
SOME OFTHESE LONGSTANDING QUESTIONSXSOMEDAYTHESEA
NSWERSMAYSEXEMASOBVIOUSTOUSASTHEXEARHTORBITING
THESUNORPERHAPSASRIDICULOUSASATOWEROFTORTOISESO
NLYTIMEWHATEVERTHATMAYBEWILXLTELXLASLONGAGOAS
THREXEHUNDREDANDFOURTYBCTHEGREXEKPHILOSOPHERA
RISTOTLEINHISBOXOKONTHEHEAVENSWASABLETOPUTFORW
ARDTWOXODARGUMENTSFORBELIEVINGTHATXTHEXEARHT
HWASAROUNDSPHERERATHERTHANA HATPLATEFIRSTTHEREAL
IZEDTHATECLIPSESOFTHEMOXONWERECAUSEDBYTHEXEARHT
HCOMINGBETWEXENTHESUNANDTHEMOXONTHEXEARHTSXSH
ADOWONTHEMOXONWASALWAYSROUNDWHICHWOULDBETRUE
ONLYIFTHEXEARHTWASXSFERICALIFTHEXEARHTXHADBEX
ENAFLATDISKTHESHADOWXWOULDHAVEBEXENELONGATEDA
NDELXLIPTICALUNLESXSTHEXECLIPSEALWAYSOCXCURXRED
ATATIMEWHENTHESUNWASDIRECTLYUNDERTHECENTEROFT
HEDISKSECONDTHEGREXEKSKNEWFROMTHEIRTRAVELSTHAT
XTHENORTHSTARAPXPEAREDLOWERINTHESKYWHENVIEWED

IN THE SOUTH THAN IT DID IN MORE NORTHERLY REGIONS SINCE THE NORTH STAR LIES OVER THE NORTH POLE IT APPEARS TO BE DIRECTLY ABOVE AN OBSERVER AT THE NORTH POLE BUT TO SOMEONE LOOKING FROM THE EQUATOR IT APPEARS TO LIE JUST AT THE HORIZON FROM THE DIFFERENCE IN THE APPARENT POSITION OF THE NORTH STAR IN EGYPT AND GREECE ARISTOTLE EVEN QUOTED AN ESTIMATE THAT THE DISTANCE AROUND THE EARTH WAS FOUR HUNDRED THOUSAND STADIA IT IS NOT KNOWN EXACTLY WHAT LENGTH A STADIUM WAS BUT IT MAY HAVE BEEN ABOUT TWO HUNDRED YARDS WHICH WOULD MAKE ARISTOTLE'S ESTIMATE ABOUT TWICE THE CURRENTLY ACCEPTED FIGURE THE GREEKS EVEN HAD A THIRD ARGUMENT THAT THE EARTH MUST BE ROUND FOR WHY ELSE DOES ONE FIRST SEE THE SAIL OF A SHIP COMING OVER THE HORIZON AND ONLY LATER SEE THE HULL? ARISTOTLE THOUGHT THE EARTH WAS STATIONARY AND THAT THE SUN, THE MOON, THE PLANETS AND THE STARS MOVE IN CIRCULAR ORBITS ABOUT THE EARTH. HE BELIEVED THIS BECAUSE HE FELT FOR MYSTICAL REASONS THAT THE EARTH WAS THE CENTER OF THE UNIVERSE AND THAT CIRCULAR MOTION WAS THE MOST PERFECT. THIS IDEA WAS ELABORATED BY PTOLEMY IN THE SECOND CENTURY AD INTO A COMPLETE COSMOLOGICAL MODEL. THE EARTH STOOD AT THE CENTER SURROUNDED BY EIGHT SPHERES THAT CARRIED THE MOON, THE SUN, THE STARS AND THE FIVE PLANETS KNOWN AT THE TIME: MERCURY, VENUS, MARS, JUPITER AND SATURN.

2 ADFGVX

2.1 De opdracht

Om ADFGVX te kraken moeten we eerst de morse code decoderen, daarna de kolom transpositie ongedaan maken en vervolgens het vierkant vinden om de oorspronkelijke tekst te bekomen.

2.2 Morse decoderen

Dit is snel gedaan door de tekst op te splitsen in de strings die tussen " / " staan en deze via een simpele map om te zetten naar hun bijbehorende characters.

2.3 De kolomtranspositie

Aangezien ADFGVX eindigt met een enkele kolomtranspositie, moesten we deze eerst ongedaan maken. Hiervoor gebruikten we hetzelfde systeem als

beschreven bij Vigenre: voor alle mogelijke transposities tot een bepaald aantal kolommen werd de index of coincidence berekend, waarna die met de hoogte index gekozen werd. Merk op dat we hierbij de digrammen moesten gebruiken om de index te berekenen. Uiteindelijk vonden we 4 mogelijke kolomtransposities met 4 kolommen, met elk een even grote IC.

2.4 Vinden van de taal.

Door eerst een frequentieanalyse te doen op de 4 meest waarschijnlijke teksten, kregen we de onderstaande tabel.

'FV'	16.93	'AA'	8.26	'DX'	8.02	'AD'	7.80
'XV'	7.32	'VD'	7.18	'AF'	6.15	'FF'	5.80
'GG'	5.64	'DD'	4.91	'XD'	3.64	'XG'	3.18
'FA'	3.10	'DG'	2.94	'VX'	1.59	'DF'	1.21
'VA'	1.16	'VG'	1.16	'GF'	0.99	'GD'	0.67
'AV'	0.56	'GX'	0.51	'FG'	0.32	'XX'	0.13
'AG'	0.13	'VV'	0.10	'FD'	0.08	'XF'	0.08
'FX'	0.05	'XA'	0.05	'DA'	0.05	'VF'	0.05
'DV'	0.05	'AX'	0.02	'GA'	0.0	'GV'	0.0

Om deze tabel te analyseren vergelijken we hem gewoon met gekende frequenties⁷. Dit kan omdat de digrammen in ADFGVX voor enkele characters staan. In ADFGVX kan men ook cijfers gebruiken. Deze staan niet in de frequentietabellen vermeld, maar dit is niet echt een probleem, aangezien deze waarschijnlijk een redelijk kleine frequentie hebben. Dit gaf ons wel een vervormde index of coincidence, waardoor we puur op deze waarde de taal niet konden bepalen. Het viel ons wel op dat het meest voorkomende digram, 'FV' bijna dubbel zo frequent was als het tweede. Van de waarschijnlijke talen voor deze tekst, zijn het Frans en Duits de enige met dit kenmerk. Aangezien de Enigma code in het Duits is, is het zeer waarschijnlijk dat deze tekst in het Frans is geschreven.

2.5 Vinden van de bron tekst

Voordat we begonnen met letters te vervangen hebben we eerst het aantal mogelijke teksten terug gebracht van 4 naar 2. We hebben dit gedaan door te kijken naar de digrammen die 2 maal op elkaar volgen. Bij twee van de vier mogelijke teksten kwam een opeenvolging van twee keer hetzelfde digram erg veel voor. Gezien dit in het Frans minder voorkomt, konden we deze teksten al elimineren.

Nu we nog 2 mogelijke teksten hadden, hebben we het werk opgesplitst onder Jakob en Robin, die elks een tekst probeerden te kraken.

⁷http://en.wikipedia.org/wiki/Letter_frequency

Door eerst de meest frequente letters in te vullen (FV=e, DX=n, ...) en daarna met trial en error de andere redelijk frequente characters in te vullen, verschenen er Franse woorden, of strings die erg leken op een Frans woord, hierdoor konden we ook de minder frequente letters invullen. Eens ongeveer de helft van de digrammen was vervangen door een character, konden we zinnen lezen en zo de andere characters aanvullen, eens we de eerste zin zeker wisten konden we Google aanspreken ⁸ en vonden we dat het ging om het eerste hoofdstuk van "*Vingt mille lieues sous les mers*" van Jules Verne.

Zoals reeds vermeld hadden we 2 teksten opgelost, beide gaven dezelfde tekst. Als we kijken naar de vierkanten die beiden gaven (Figure 2 en 3), zien we dat de ene de getransponeerde versie van de andere is. Dit betekent dat de digrammen voor de ene versie de omgekeerde zijn van de andere versie. Ene ene ciphertekst is dus gelijk aan de andere ciphertekst met elk digram omgedraaid, voor de kolomtranspositie. Gezien de kolomtranspositie met een even aantal kolommen gebeurde, kan door een verschillende kolomtranspositie toe te passen op beide teksten, dezelfde uiteindelijke ciphertekst bekomen worden.

	A	D	F	G	V	X
A	a	5	c	?	q	w
D	s	o	2	x	i	d
F	r	g	l	b	0	1
G	6	m	y	u	f	p
V	h	3	e	?	z	t
X	4	n	8	j	v	k

Figure 2: Vierkant van de 1e tekst

	A	D	F	G	V	X
A	a	s	r	6	h	4
D	5	o	g	m	3	n
F	c	2	i	y	e	8
G	?	x	b	u	?	j
V	q	l	0	f	z	v
X	w	d	1	p	t	k

Figure 3: Vierkant van de 2e tekst

3 Enigma

3.1 Opdracht

De vierde ciphertekst was versleuteld met Enigma. Eerst implementeerden we een volledige Enigma machine. Daarna gingen we op zoek naar de gebruikte rotoren en hun beginstand.

⁸https://www.google.be/search?q=1%27annee+1966+fut+marquee+par+une+evenement&oq=1%27annee+1966+fut+marquee+par+une+evenement&aqs=chrome..69i57j1740j0j7&sourceid=chrome&es_sm=122&ie=UTF-8#q=1%27annee+1966+fut+marquee+par+un+evenement+bizarre

3.2 Een beetje extra onderzoek

Voor we aan het kraken van enigma begonnen hebben we eerst het internet nog gebruikt om wat extra informatie te verzamelen. Deze info en de nota's van de les stelden ons in staat om de tekst relatief eenvoudig te kraken. Hieronder een korte oplijsting met de belangrijkste bronnen van informatie die we gebruikten om meer te leren over Enigma, de geschiedenis en het kraken.

- 2 videos van Numberphile op Youtube over enigma:
 - https://www.youtube.com/watch?v=G2_Q9FoD-oQ
Een video over de werking van Enigma, de internals, het elektrisch circuit ed.
 - <https://www.youtube.com/watch?v=V4V2bpZlqx8>
Een video over hoe de Engelsen de code konden kraken. Deze was niet echt van nut voor het kraken dat wij moesten doen, omdat wij reeds een Crib gekregen hadden, dit gaat meer over hoe de engelsen zon crib konden vinden.
- 2 Videos van Computerphile over Bletchley park en Enigma
 - https://www.youtube.com/watch?v=d2NWPG2gB_A https://www.youtube.com/watch?v=kj_7Jc1mS9k Een iets meer in depth gesprek over de geschiedenis van Enigma en de handelingen in Bletchley park.
- De wikipedia pagina's over Enigma en Alan Turing
 - http://en.wikipedia.org/wiki/Alan_Turing
 - http://en.wikipedia.org/wiki/Enigma_machine
- <http://www.bletchleypark.org.uk/content/hist/>
De Pagina van Bletchley Park met een heel stuk geschiedenis rond Enigma.

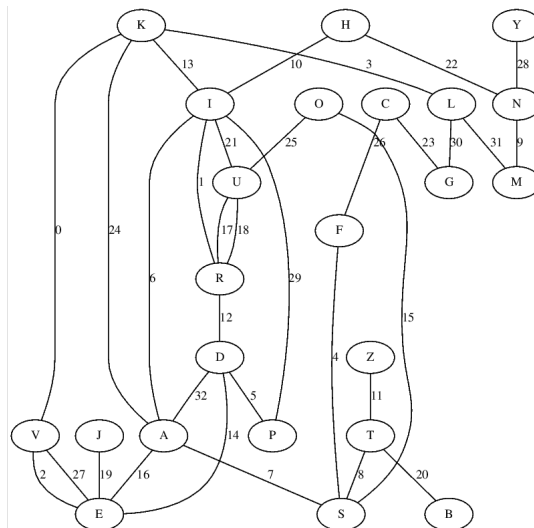
3.3 Maken van de Enigma machine

Eerste stap om enigma te kunnen ontcijferen was om een werkende enigma machine te maken. Als taal kozen we voor Go, omdat deze het heel eenvoudig maakt om concurrent functies uit te voeren. Op deze manier zouden we alle rotor posities tegelijk kunnen proberen, net als "The Bombe" dit deed in wereldoorlog II. De implementatie was niet zo moeilijk, met de info die we hadden uit de les, konden we eenvoudig de verschillende componenten implementeren: Rotor, Reflector en Plugboard. Voor elk van deze componenten schreven we testen zodat we zeker waren dat er geen fouten inzaten

voor we ze samenbrachten in de Enigma klasse. Deze zorgde voor het linken van de componenten: gaf character per character aan de componenten en concateneerde de ze terug tot de geëncrypteerde of decrypteerde tekst. Om de volledige machine eenvoudig te kunnen gebruiken hebben we een JSON parser toegevoegd die een JSON file neemt en uit de info in deze file een enigma construeert. Een voorbeeld file kan u vinden in de appendix. Om de machine te testen encrypteerden we enkele teksten via de webapp⁹ die de professor ons had bezorgd en lieten ze dan door onze machine decoderen.

3.4 Crib graph en begin positie van de rotors

Er was een crib gegeven, dus we konden makkelijk een crib graph opstellen. Hiervoor stelden we manueel een .dot file om, waarmee we een visuele voorstelling genereerden. Deze is hieronder weergegeven.



Aan de hand van de graph konden we op zoek naar gesloten paden. Door de lengte van de crib waren er heel wat mogelijkheden. We bepaalden voor de letter U 6 gesloten paden. We gingen dan, met elke mogelijke volgorde van rotoren, op zoek naar een k waarvoor een letter invariant bleef op elk van die paden. Hiervoor was een kleine modificatie aan onze geïmplementeerde Enigma machine voldoende. We vonden dat er hiervoor slechts één optie was, namelijk rotorvolgorde 420 met als beginstand KSY. De invariante letter was hier de U, wat betekent dat de U door het plugboard niet aangepast wordt. We herhaalden deze test met de letters A, D en I, waarvoor we ook telkens 5 à 6 gesloten paden zochten. Hier vonden we ook telkens slechts 1 of 2 resultaten, waaronder altijd **rotorvolgorde 420 met beginstand KSY**. Dit was dus zeker het juiste antwoord. Uit deze resultaten vonden

⁹<http://www.algebra.ua.ac.be/stijn/enigma.phtml>

we ook dat het plugboard I en Y verwisselt, en de A en de D niet aanpast. Nu moesten we enkel nog op zoek naar de rest van het plugboard.

3.5 Bepalen van het plugboard en ontcijferen van de tekst

Omdat we een behoorlijk grote crib hadden, waren er voor de meeste letters gesloten paden te vinden in de graph. We konden dus gewoon een versimpelde versie van de voorgaande code draaien voor elke letter met gesloten paden om het plugboard verder te bepalen. Hiervoor hoefden we de test enkel voor de correcte rotorvolgorde en beginstand draaien. De enige letters zonder gesloten paden waren B, J, Q, T, W, X, Y en Z waarbij Q, W en X helemaal niet in de graph voorkwamen, en we al wisten dat Y met I werd omgewisseld. We gingen dus enkel de mappings van de letters met gesloten paden zoeken. Hierdoor vonden we ook de mappings voor B, Q, W en Z omdat ze elk verwisseld werden met een letter die wel gesloten paden had. We hadden het plugboard dus op J, T en X na bepaald. Er waren nu echter maar 4 mogelijke plugboards over: ofwel werden twee van de drie letters met elkaar gewisseld ofwel werd geen enkele gewisseld. Door de ciphertekst met elk van de vier opties te decipheren, vonden we dat enkel die waarbij J, T en X niet werden aangepast de crib juist ontcijferde en dus juist was. Hiermee hadden we ook het plugboard volledig bepaald en konden we de tekst volledig ontcijferen. Het was een deel uit Die Blechtrommel van Günter Grass ¹⁰.

Rotorvolgorde: 420

Beginstand: KSY

Plugboard: KEDCHGFYJBLWNZPSRQTUVMXIO

Verwisselingen in plugboard: B-K, C-E, F-H, I-Y, M-W, O-Z, Q-S

VIELSPASSMITDIESERUEBUNGAUFENIGMAZUGEGBENICHBIN
INSASSEINERHEILUNDPFLEGEANSTALTMEINPFLEGERBEOBA
CHTETMICHLASSTMICHKAUMAUSDEMAUGEDENNINDERTURIS
TEINGUCKLOCHUNDMEINESPFLEGERSAUGEISTVONJENEMBRA
UNWELCHESMICHDENBLAU AUGIGENNICHTDURCHSCHAUENKA
NNMEINPFLEGERKANNALSOGARNICHTMEINFEINDSEINLIEBGE
WONNENHABEICHIHNERZAHLEDEM GUCKERHINTERDERTURSO
BALDERMEINZIMMERBETRITTBEGEBENHEITENAUSMEINEMLE
BENDAMITERMICH TROTZDESIHNHINDERNDENGUCKLOCHESKE
NNENLERNTDERGUTESCHEINTMEINEERZAHLUNGENZUSCHAT
ZENDENNSOALDICH IHMETWASVORGELOGENHABEZEIGTERM
IRUMSICHERKENNTLICHZUGEBENSEINNEUESTESKNOTENGEBI
LDEOBEREINKUNSTLERISTBLEIBEDAHINGESTELLTEINEAUSST
ELLUNGSEINERKREATIONENWURDEJEDOCHVONDERPRESSEG
UTAUFGENOMMENWERDENAUACHEINIGEKAUFERHERBEILOCK

¹⁰<http://www.lawrenceglatz.com/germ3230/texte/grass1.htm>

ENERKNOTETORDINAREBINDFADENDIEERNACHDENBESUCHSS
TUNDENINDENZIMMERNSEINERPATIENTENSAMMELTUNDENT
WIRRTZUVIELSCHICHTIGVERKNORPELTENGESPENSTERNTAU
CHTDIESEDANNINGIPSLASSTSIEERSTARRENUNDSPIESSTSIEMIT
STRICKNADELNDIEAUFHOLZSOCKELCHENBEFESTIGTSINDOFT
SPIELTERMITDEMGEDANKENSEINEWERKEFARBIGZUGESTALT
ENICHRATEDAVONABWEISEAUFMEINWEISSLACKIERTESMETA
LLBETTHINUNDBITTEIHNSICHDIESESVOLLKOMMENSTEBETT
UNTBEMALTVORZUSTELLENENTSETZTSCHLAGTERDANNSEINE
PFLEGERHANDEUBERDEMKOPFZUSAMMENVERSUCHTINETWA
SZUSTARREMGESICHTALLENSCHRECKENGLEICHZEITIGAUSDR
UCKZUGEBENUNDNIMMTABSTANDVONSEINENFARBIGENPLAN
ENMEINWEISSLACKIERTESMETALLENESANSTALTSBETTISTAL
SOEINMASSSTABMIRISTESSOGARMEHRMEINBETTISTDASENDLI
CHERREICHTEZIELMEINTROSTISTESUNDKONNTEMEINGLAUBE
WERDENWENN MIRDIEANSTALTSLEITUNGERLAUBTEEINIGEAN
DERUNGENVORZUNEHMENDASBETTGITTERMOCHTEICHERHO
HENLASSENDAMITMIRNIEMANDMEHRZUNAHETRITTEINMALIN
DERWOCHEUNTERBRICHTEINBESUCHSTAGMEINEZWISCHENW
EISSENMETALLSTABENGEFLOCHTENESTILLEDANNKOMMENS
IEDIEMICHRETTENWOLLENDENENESSPASSMACHTMICHZULIEB
ENDIESICHINMIRSCHATZENACHTENUNDKENNENLERNENMOCH
TENWIEBLINDNERVOSWIEUNERZOGENSIESINDKRATZENMITIH
REN FINGERNAGELSCHERENANMEINEMWEISSLACKIERTENBET
TGITTERKRITZELNMITIHRENKUGELSCHREIBERNUNDBLAUSTI
FTENDEMLADELANGGEZOGENEUNANSTANDIGESTRICHMANNC
HENMEINANWALTSTULPTJEDESMALSOBALDERMITSEINEMHAL
LODASZIMMERSPRENGTDENNYLONHUTUBERDENLINKENPFOS
TENAMFUSSENDEMEINESBETTESSOLANGESEINBESUCHWAHRT
UNDANWALTEWISSENVIELZUERZAHLENRAUBTERMIRDURCHD
IESENGEWALTAKTDASGLEICHGEWICHTUNDDIEHEITERKEITN
ACHDEMMEINEBESUCHERIHREGESCHENKEAUFDEMWEISSEN
MITWACHSTUCHBEZOGENENTISCHCHENUNTERDEMANEMONEN
AQUARELLDEPONIERTHABENNACHDEMESIHNENGELUNGENIST
MIRIHREGERADELAUFENDENODERGEPLANTENRETTUNGSVER
SUCHEZUUNTERBREITENUNDMICHDENSIEUNERMUDLICHRETT
ENWOLLENVOMHOHENSTANDARDIHRERNACHSTENLIEBEZUUB
ERZUEGENFINDENSIEWIEDERSPASSANDEREIGENENEXISTENZU
NDVERLASSEN MICHDANNKOMMTMEINPFLEGERUMZULUFTE
UNDDIEBINDFADENDERGESCHENKPACKUNGENEINZUSAMMEL
NOFTMALSFINDETERNACHDEMLUFTENOCHZEITANMEINEMB
ETTSITZENDBINDFADENAUFDROSELNDSOLANGESTILLEZUVER
BREITENBISICHDIESTILLEBRUNOUNDBRUNODIESTILLENENNEB
RUNOMUNSTERBERGICHMEINEJETZTMEINENPFLEGERLASED

ASWORTSPIELHINTERMIRKAUFTEAUFMEINERECHNUNGFUNFH
UNDERTBLATTSCREIBPAPIERBRUNODERUNVERHEIRATETKI
NDERLOSISTUNDAUSDEMSAUERLANDSTAMMTWIRDSOLLTEDE
RVORRATNICHTREICHENDIEKLEINESCHREIBWARENHANDLUN
GINDERAUCHKINDERSPIELZEUGVERKAUFTWIRDNOCHEINMAL
AUFsuchenUNDMIRDENNOTWENDIGENUNLINIERTENPLATZFU
RMEINHOFFENTLICHGENAUESERINNERUNGSVERMOGENBESCH
AFFENNIEMALSHATTEICHMEINEBESUCHERETWADENANWALT
ODERKLEPPUMDIESENDIENSTBITTENKONNENBESORGTEMIRV
ERORDNETELIEBEHATTEDENFREUNDENSICHERVERBOTENET
WASSOGEFAHRLICHESWIEUNBESCHRIEBENESPAPIERMITZUBR
INGENUNDMEINEMUNABLASSIGSILBENAUSSCHEIDENDENGEST
ZUMGEBRAUCHFREIZUGEBENALSICHZUBRUNOSAGTEACHBRU
NOWURDESTDUMIRFUNFHUNDERTBLATTUNSCHULDIGESPAPI
ERKAUFENANTWORTETEBRUNOZURZIMMERDECKEBLICKEND
UNDSEINENZEIGEFINGEREINENVERGLEICHHERAUSFORDERND
INDIEGLEICHERICHTUNGSSCHICKENDSIEMEINENWEISSESPAPIE
RHERROSKAR