

**FOM - Hochschule für Oekonomie & Management
Hamburg**

**Master-Studiengang Big Data & Business Analytics
2. Semester**

**Development of a solution for genetic analysis of
ALL genomes by implementing Latent Dirichlet allocation**

Betreuer: Prof. Dr. Martin Münstermann

Autor: Jacqueline Franßen

Matrikel-Nr: 496804

2. Fachsemester

Hamburg, den 25.07.2019

Contents

1	Abstract	2
2	Introduction	3
3	Related work	4
4	Latent Dirichlet Allocation (LDA)	7
4.1	General description	7
4.2	Examples and possible use cases	11
4.3	Python library 'Gensim'	12
5	Acute Lymphoblastic Leukemia	13
5.1	Types of Leukemia and its causes	13
5.2	Examples for Genome Analysis: Next-Generation Sequencing (NGS) .	15
5.3	Data sources: National Center for Biotechnology Information (NCBI) and Ensembl genome browser 96	16
6	Development of a solution for genetic analysis of Acute Lymphoblastic Leukemia (ALL) genomes by implementing LDA	19
6.1	Problems and challenges of genetic analysis	19
6.2	First steps: Draft of developed solution	20
6.3	Proposed solution	22
6.4	Results	26
7	Conclusion and Outlook	27
7.1	Lessons learned	27
7.2	Conclusion	27
7.3	Outlook	28

8 Abbreviations	29
9 Bibliography	36
10 Appendix A	37
11 Appendix B	38

List of Figures

5.1	NCBI query to find all ALL related gene sequences	17
5.2	Detailed information about selected gene sequence	17
5.3	Information about genome IKZF1, found in Ensembl genome browser 96	18
6.1	Extracted data (including genome's name and description), the green rectangle includes the descriptions	20
6.2	Diagram of first draft of developed algorithm	21
6.3	Diagram of developed algorithm to create basic topics among given genomes	23
6.4	Diagram (created with PyLDAvis) which shows the distribution of the topics	24
6.5	Diagram of LDA process	25
6.6	Call of method LdaModel() from library Gensim and its output	26

List of Tables

1 Abstract

There exist many types of cancer. Every type of cancer is so individual and is treated particularly in practice. Experimental studies showed that cancer is the consequence of the incorrect mutation of genomes which makes it hard to find out the appropriate solution and therapy for the patient (since every patient has its own individual set of genomes). Generally, our human genomes are nearly similar, except from small variations (due to ethnics, demography etc.). Besides, to find out which mutations can lead to a certain kind of cancer, data analysis algorithms can be a chance to find out or to 'predict' the appearing mutations.

On the other side, from a computer scientist's view, there exist many algorithms to analyze data. In this paper, the focus is on LDA, an unsupervised (autonomous) topic modelling technique. Given a certain dataset (including genomes which can cause ALL), topics and similarities between the genome data are figured out. The results are discussed in section 6.4.

The aim of this paper is to find out whether LDA is the appropriate solution to analyze genome data and to find out similarities between the genes. These similarities are expected as a solution approach to find a pattern among the disease-related genomes and to be able to predict these mutations. Furthermore, if some patient shows these genomes (or mutations of them), the geneticist or molecular biologist can inject the required sequence to the patient so that the cancer will not break out.

2 Introduction

LDA provides many advantages, such as that it is suitable for large data and that it performs very well in extracting topics of text documents ¹. In the year 2000, more than 209,000 children and adults died of leukemia globally. This number keeps growing year by year ². In order to figure out whether there is a solution of analyzing leukemia related genome data and predicting the genome's mutations, an algorithm has been developed (see code in chapter 11).

To give a short overview of this paper, it is ordered as follows: The first section 3 will explore the similar scientific articles and papers related to genome analysis, cancer and natural language processing techniques. After that, section 5 gives a short description of the disease ALL and refers to genome libraries. Consequently, section 6 focusses on the developed implementation and topic modelling algorithm. Finally, section 7 summarizes the research and implemented methods of this paper and gives an outlook for future analysis.

¹cf.[Twinandilla et al. 2018]

²cf.[Tang, Cao, and Y. Wang 2011]

3 Related work

In this section, some examples of using text mining techniques in biology and medicine are described. Zhao et al.¹ describe how topic modeling can be used to analyze NGS. Generally, by implementing topic modelling, text corpus are generated.

In the beginning of every genome analysis, there are several important questions to ask. Jurca et al.² recommend to ask the following questions: What are the top studied genes in breast cancer? What are the regulations and limitations of blood cancer research in every country? Which countries have studied the largest number of breast cancer? Which are the popular genes mentioned together by countries every year? Where do key genes lie in the soft clusters?

Jurca et al. describe a process to use large-scale text analysis of biomedical abstracts in order to generate new hypothesis about cancer biomarkers. The target is to develop a data mining methodology to find out the genes associated with cancer. By analyzing disease-specific gene expression, experimental data is being checked. The key question is whether a gene has indeed been upregulated or downregulated with respect to a disease.

According to Xu et al.³, micro Ribonucleic Acid (MIRNA)s build a class of 17-27 nucleotides single-stranded Ribonucleic Acid (RNA) molecules that regulate gene expression post-transcriptionally. In the described text-mining process, Xu et al. identified nine MIRNAs in bladder cancer and adopted protein-protein interaction sites between these miRNAs and target genes. The results of the analyzation process lead to two relationship types between bladder cancer and its MIRNA: casual and unspecified.

Topic modelling is not only used to analyze relationships between genomes but also to

¹cf.[Zhao et al. 2016]

²cf.[Jurca et al. 2016]

³cf.[Xu et al. 2013]

improve diagnoses for stroke disease. Djatna et al.⁴ describe an 'Intuitionistic Fuzzy Based Decision Tree' to diagnose different types of stroke disease. To be precise, the different types of stroke diseases can be calculated by a Hamming distance. The term 'Fuzzy logic' means logic that underlies the reasoning of data using precise estimates. It is the fastest way to map input space into output space using a degree of membership.

Lloret et al.⁵ built an automatic summarization algorithm for literature. It can includes three steps: First, topic identification, second topic interpretation and third summary generation. While describing the process of textual analyzation, Lloret et al. mention a specific term: term frequency inverse document frequency (TFIDF) which is important for topic modelling. In addition to topic-based approaches, there are graph-based approaches and discourse-based approaches. Graph-based approaches implicate nodes that represent text elements and the edges/links refer to synonymy⁶. Discourse-based approaches include Rhethorical Structure Theory (RST), Hidden-Markov-Models (HMM) or Bayesian models (BM).

Yang et al.⁷ describe the process of 'constructing a database for relations between human copy number variant (CNV)s and human genetic disease via systematic text mining'. In general, CNV can cause disease, e.g. by manipulating gene dosage, disruption, fusion or other genetic position effects. To be more precise, there can CNV can lead to two types of autosomal variants: They can either cause deletion or amplification of the long or broken arm region of chromosomes 1-22 or can build multiples of chromosomes 1-21 (e.g. as in disease trisomy 21).

According to their article, Yang et al.⁸ used a CNV database which linked the CNV information to the NCBI Gene and Ontology database. Yang et al. mention three steps in the text mining process. First, during the pre-processing step, unstructured fields are split into separated sentences by using Natural Language Toolkit (NLTK), a python package⁹. After that, in the named entity recognition (NER) step, all disease mentions within the DNorm system, such as MeSH IDs are recognized. In the third

⁴cf.[Djatna, Hardhienata, and Masruriyah 2018]

⁵cf.[Lloret and Palomar 2012]

⁶cf.[Lloret and Palomar 2012]

⁷cf.[Yang et al. 2018]

⁸cf.[Yang et al. 2018]

⁹cf.[*Natural Language Toolkit — NLTK 3.4.1 documentation* 2019]

step, Relation extraction (RE), the positions in sentences and entities are compared to generate instances that consist of two candidate entities within one single sentence. Yang et al. mention two more processing methods: Parallel Processing and Post Processing which includes data cleaning and statistics. The term 'data cleaning' is explained as 'de-duplicating data after each step of the process to reduce repetitive operations and prevent statistical errors'. This is a very useful step in biomedicine since biomedical databases may contain errors. For that reason, users can give feedback through a feedback mechanism to improve the quality of the databases.

Lu et al.¹⁰ used multi-channel LDA to model healthcare data. In fact, by creating a learned latent variable model, the likelihood of a set of diagnosis, medications, contextual information in a patient's record can be evaluated. This can help to identify outliers and improve medical data quality. Furthermore, disease groups can be identified, or missing medication or diagnosis can be predicted. Lu et al. used Association Rule Mining (ARM) and supervised learning to predict missing medications. The topic model defines how words in a document are generated through the control of latent topics.

¹⁰cf.[Lu, Wei, and Hsiao 2016]

4 LDA

4.1 General description

LDA was developed by David Blei et. al in the year 2003 and is a clustering algorithm for text mining. It counts to the most popular topic modelling algorithms¹. According to Zhao et al., topic modelling requires of a number documents which represent each of them a mixture of latent topics. Moreover, each topic is expressed by a distribution of words. During LDA, two relationships are analyzed: First, the relationship between documents and words, also called 'per-document topic distributions'. Second, the relationship between words and topics ('per-topic word distributions'). To measure the relationships exactly and to make inference about topics and documents for text mining, probability matrices are calculated.

Park et al.² define topic models as follows: Documents are no longer a collection of words, but a collection of topics. Furthermore, LDA is a generative topic model which uses a dirichlet parameter (also called dirichlet prior) to model documents. By changing the dirichlet prior, the number of topics that the model assigns to each word and document can be controlled. To be more precisely, a small dirichlet prior means a small number of topics assigned to each word. By increasing the dirichlet prior, the distribution of topics to each word rises. Moreover, the dirichlet parameter is obtained for each document and can be fitted using a maximum or estimated likelihood. If the dirichlet parameter does not fit, the gain in computational efficiency is obtained. Otherwise, there is no advantage (when dirichlet parameter fits well).

According to Jurca et al.³ the text mining process can be divided into four steps: First, the information has to be retrieved by user queries (Information Retrieval (IR)).

¹cf.[Zhao et al. 2016]

²cf.[L. A. F. Park and Ramamohanarao 2009]

³cf.[Jurca et al. 2016]

Second, different vocabularies and ontologies have to be integrated (NER). Third, during Information extraction (IE), relationships between biological entities in the texts are extracted by either using co-occurrence processing or Natural Language Processing (NLP). Last, there has to be gained biologically meaningful knowledge about how biological entities are related by implementing Knowledge Discovery (KD) methods. Moreover, there can be distinguished between three types of clustering: hard clustering, hierarchical clustering and soft clustering. Hard clustering describes the process of separating items into distinct groups where each item is exactly in one cluster. Hierarchical clustering implicates creating single-link clusters (how similar the items are to one another) and complete-link (how dissimilar the items are). Soft clustering means that items cannot be distinctly separated into clusters and partly are member of two or more clusters at a time.

Besides, Djatna et al.⁴ mention data mining techniques, such as Classification and Regression Tree (CART), Iterative Dichotomized 3 (ID3), Decision Tree (DT), Principal Component Analysis (PCA) and LDA.

Lu et al.⁵ define topic models as a text mining approach that assumes observed word co-occurrences which are governed by latent variables. LDA includes the identification of latent topics from a set of documents, analyzing long-term topic trends and modelling words and references in documents.

According to Hoffman et al., LDA is a probabilistic (Bayesian) model of text documents⁶. The idea of LDA is to define a document as a collection of k topics. Each topic defines a multinomial distribution over a vocabulary which is drawn from a dirichlet.

What is more, every term has a probabilistic relationship to every document. The topic model probabilities are stored as term relationships in thesaurus. The term frequencies are stored in the document index.

Twinandilla et al.⁷ mention three variables to be defined before the LDA process: α (the diversity of sentence distribution), β (the diversity of topic distribution) and γ (the similarity between sentences and titles).

⁴cf.[Djatna, Hardhienata, and Masruriyah 2018]

⁵cf.[Lu, Wei, and Hsiao 2016]

⁶cf.[Hoffman, Bach, and Blei 2010]

⁷cf.[Twinandilla et al. 2018]

In their article 'Multi-document summarization using k-means and LDA-significant sentences, Twinandilla et al. ⁸ describe the research process by implementing the following six steps.

1.Step: Preprocessing

First, all words and sentences need to be simplified by using a bag of words as well a bag of sentences. In detail, this step includes case folding (putting all words into lower case), tokenization (cutting a document into an array of words or sentences and eliminating punctuation), stopword removal (deleting words that appear often without a particular meaning) and stemming (changing words in a document that appear often without a particular meaning).

2.Step: Calculate the number of clusters

Second, the number of clusters needs to be calculated by using k-means clustering.

3.Step: LDA

In this step, Twinandilla et al. distinguish between generative and inference LDA. Generative LDA forms a document from a collection of words whereas inference LDA only retrieves information from documents.

4.Step: Sentence LDA

Fourth, during sentence LDA, documents are represented as topic representation. Each topic is a sentence distribution that represents a sentence. This sentence has significant weight on a multi-document summarization.

5.Step: Summary formation

In this step, each document is sorted by a decreasing value of the final sentence weight. After that, the p percent of sentences with the highest value has to be chosen from each document. The p value is subsequently called 'summarization level'.

6.Step: Arrange selected sentences in a sequence

The last step includes the arrangement of all selected sentences in a sequence. This means putting them into a useful order to summarize all documents.

⁸cf.[Twinandilla et al. 2018]

As reported by Blei et al.⁹, LDA is a generative probabilistic model for collections of discrete data such as text corpora. In addition to that, LDA is represented as three level hierarchical Bayesian model in which each item of a collection is a finite mixture over an underlying set of topics. What is more, each topic is modeled as an infinite mixture over an underlying set of probabilities. Blei et al. define topic probabilities as an explicit representation of a document.

TFIDF TFIDF is a scheme through which a basic vocabulary of words or terms is chosen. For each document in the corpus a count (which represents the number of occurrences for each word) is formed¹⁰. There are three terms to be distinguished: First, a word is a basic unit of discrete data, defined to be an item from the vocabulary indexed by $1 \dots V$. Second, a document is a sequence of N words. Third, a corpus is a collection of M documents. After a suitable normalization process, the term frequency count is compared to an inverse document frequency count. This leads to the total number of occurrences of a word in the entire corpus. The result is a term-by-document matrix X whose columns contain TFIDF values for each document in the corpus.

LDA process The process of LDA can be briefly described with the following step¹¹: First, there are M documents in the corpus. Second, each document j has N_j words. In the next step, the observed value w_{ji} describes the appearance of a word i in a document j . In addition to that, all words will be clustered into K topics which are defined as object classes. Finally, each topic k is modelled as a multinomial distribution over the codebook.

As reported by Wang et al., LDA is a language model which clusters co-occurring words into topics. Moreover, documents are described as 'bag of words'. Wang et al. describe a special form of LDA: Spatial LDA. It encodes spatial structure among visual words, assuming the partition of words into documents is known a priori.

⁹cf.[Blei, Ng, and Jordan 2003]

¹⁰cf.[Blei, Ng, and Jordan 2003]

¹¹cf.[X. Wang and Grimson 2008]

LDA extensions Extensions of LDA are author-topic model, dynamic-topic model and correlated topic model. Wang et al. refer to the dynamic-topic model while describing how visual words are clustered into topics which correspond to object classes.

4.2 Examples and possible use cases

Zhao et al. describe the process of analyzing genomes as follows: First, each document corresponds to one of the total number of Desoxyribonucleic acid (DNA) strains. Second, all documents had the same number of words. Third, the distribution of words for topics as well as the distribution of topics in documents were described by random variables obeying Dirichlet distributions with parameters α and β . After that, nucleotides and their orders in NGS sequences could be treated as words and the genetic information in sequences was translated and exhibited as a 'bag of words'¹². By using the strain-topic matrix derived from topic modelling, relationships or similarities between the strains serotypes can be found out.

Hoffman et al.¹³ describe the development of an online variational Bayes algorithm for LDA which is based on stochastic optimization with a natural gradient step. This step converges to a local optimum of the variational Bayes objective function. To be more precise, Bayesian models provide a natural way to encode assumptions about observed data. There can be distinguished between two approaches: First, sampling approaches are based on Markov Chain Monte Carlo (MCMC) sampling. Here, a Markov chain defines the stationary distribution. Second, there are optimization approaches which are usually based on variational inference MCMC. In this case, variational Bayes optimizes the simplified parametric distribution.

Twinandilla et al.¹⁴ developed a 'multi-document summarization using k-means and LDA-significant sentences' on yellow journalism. The term 'yellow journalism' stands for 'redundant news documents' which makes it difficult to distinguish documents containing fact or opinionated information. After defining the corpus, Twinandilla et al. describe two different summarization processes: Abstractive as well as extractive summarization. Abstractive summarization means summarizing documents by creat-

¹²cf.[Zhao et al. 2016]

¹³cf.[Hoffman, Bach, and Blei 2010]

¹⁴cf.[Twinandilla et al. 2018]

ing new sentences (with the same information as the original document). Extractive summarization suggests summarizing a document by selecting a part of a sentence in that document.

As described in chapter 3 on page 4, Lu et al.¹⁵ used Multiple-channel Latent Dirichlet Allocation (MCLDA) to estimate latent health status groups. MCLDA constructs latent relations among diagnoses, medications, contextual variables in different status groups. To be more precise, it is a dimensional reduction method that summarizes each record using a probability vector over a latent health status group. The prediction tasks were performed by using a Collapsed Gibbs model (CGS) based inference model and inferred methods. During the topic modelling process, Lu et al. refer to two associations among the data: diagnosis-medication associations to identify the clinical use of medications and diagnosis-diagnosis associations to create a network structure among the diseases.

Lee et al.¹⁶ analyzed the direct and indirect interactions between DNA copy number and gene expression changes. 'DNA copy number aberrations and gene expression changes provide valuable information for studying chromosomal instability and its consequences in cancer.' Three types of cancer have been analyzed: brain, bladder and breast cancer. The results of the analysis lead to a significant association between DNA copy number and its gene expression. By monitoring gene expression data, different classes of tumors can be distinguished successfully.

4.3 Python library 'Gensim'

'Gensim' stands for 'generate similar' and is a python library, created by Radim Rehurek and Petr Sojka¹⁷. Firstly created in 2008, it was a collection of various Python scripts to generate a short list of the most similar articles to a given article. Gensim supports scalable statistical semantics to analyze plain-text documents for semantic structure and to retrieve semantically, similar documents.

¹⁵cf.[Lu, Wei, and Hsiao 2016]

¹⁶cf.[Lee, Kong, and P. J. Park 2008]

¹⁷cf.[Řehůřek and Sojka 2010]

5 Acute Lymphoblastic Leukemia

5.1 Types of Leukemia and its causes

According to Jurca et al.¹, cancer is the result of damage, especially of mutations to cell's DNA which leads to a cell losing its normal functionality and gains the ability to indefinitely multiply until normal tissue functions are impaired. This is also why malignant cancer is distributing so fast. Besides, each patient develops a different set of cancerous mutations in various genes which lead to multiple subtypes of cancer. Furthermore, some genes can be up-regulated (which means that they are transcribed and expressed more), down-regulated (which means that they are not expressed) or can be co-expressed (which means that they are expressed at the same time).

As stated by Montaña et al., ALL is a malignant disorder originating from hematopoietic B-/T-cell precursors which are characterized by marked heterogeneity at molecular and clinical levels². There are many approaches to analyze these precursors, such as analyzing targeting of transcriptional factors (PAX5) which are involved in the pathogenesis of B-ALL. Other therapeutic and clinical approaches are genome editing techniques, i.e. the design of new therapies (Chimeric Antigen Receptors (CAR)s) and the study of genes involved in the evolution of pathogenesis.

Jagadev et al.³ define leukemia as blood cancer which begins in the bone marrow where it causes formation of a large number of abnormal cells. According to Jagadev et al. there are four common types of leukemia: ALL, Acute Myeloid Leukemia (AML), Chronic Lymphocytic Leukemia (CLL), Chronic Myeloid Leukemia (CML). Furthermore, the normal and leukemic lymphocytes differ significantly.

¹cf.[Jurca et al. 2016]

²cf.[Montaña et al. 2018]

³cf.[Jagadev and Virani 2017]

As explained by Tang et al.⁴, every subtype of leukemia can have individualized treatments and therapies.

According to Andersson et al., ALL is caused the expansion of immature hematopoietic cells (blasts) in the bone marrow and peripheral blood⁵. Furthermore, ALL counts to the most common childhood malignancies, appearing between 2 to 5 years of age. ALL has two subtypes: T-Acute Lymphoblastic Leukemia (T-ALL) and B-cell precursor Acute Lymphoblastic Leukemia (BCP-ALL). The new cases of T-ALL are 15% childhood and 25% adult ALL. Compared to BCP-ALL, T-ALL is characterized by older ages of onset, male sex predominance, inferior outcome and chromosomal abnormalities. In 1960, the 'Philadelphia chromosome' in CML was discovered which groups it to 'genetic disease'. The result of the exchange of genetic material between two chromosomes is called chromosomal translocations. These serve as hallmarks of leukemia and are intimately associated with a specific leukemia type and its prognosis. Moreover, the different types of ALL are commonly associated with genetic events, such as DNA copy-number alterations (CNA), a sub-microscopic deletions or gains, amplifications and sequence mutations.

Therapies against ALL The Childhood Acute Lymphoblastic Leukemia Collaborative Group (CALLCG) first met in 1994 to find out a specific treatment for childhood ALL. In 2009, CALLCG describe a therapy using anthracyclines to treat childhood ALL⁶. In fact, some studies suggest that anthracyclines might cause cardiotoxicity. In most cases of the therapeutic approach, anthracyclines significantly reduced the bone marrow relapse when added to the standard therapy. But it did not significantly increase an event-free survival. One year after that, CALLCG reviewed the addition of vincristine plus steroid pulses (prednisone and prednisolone pulses) to the treatment of childhood ALL⁷. As a result, the combination of vincristine pulses significantly reduced the overall event rate by around 30% to 10% improvement in event-free survival (EFS) by five years.

⁴cf.[Tang, Cao, and Y. Wang 2011]

⁵cf.[Andersson et al. 2016]

⁶cf.[Group (CALLCG) 2009]

⁷cf.[“Systematic review of the addition of vincristine plus steroid pulses in maintenance treatment for childhood acute lymphoblastic leukaemia – an individual patient data meta-analysis involving 5659 children” 2010]

5.2 Examples for Genome Analysis: NGS

NGS refers to post-Sanger sequencing methods⁸. Since NGS produces large volumes of sequence data it might be very useful implementing topic modelling techniques in order to maintain the flexibility for the level of resolution required for given experiments. According to Gasperskaja et al.⁹, NGS does not require a priori knowledge about genomic feature, it only requires a low amount of DNA or RNA as input.

The step before analyzing two or more (multiple) genomes is called alignment which includes a comparison of two genomes. There are many different types of alignments, but Zhao et al. refer to the Multiple Sequence Alignment (MSA) by describing Multiple Sequence Comparison by Log- Expectation (MUSCLE) and CLUSTAL.

Gasperskaja et al. mention an important question which should be asked before every genome analysis: 'Is the variance pathogenic?' and whether there is any relationship between genotype and phenotype which means that it can lead to a disease or can cause a number of disorders. Moreover, there can be distinguished between beneficial (Single Nucleotide Polymorphism (SNP)) and pathogenic (nonsense variant) single nucleotide changes, large microscopically visible or chromosomal aberration. To find out whether a genome mutation is pathogenic, Gasperskaja et al. explain that substantial information about functional genomics can be found through the analysis of messenger RNA (mRNA) or complementary Ribonucleic Acid (cRNA) (which is a copy from mRNA by reverse transcription Polymerase Chain Reaction (PCR)). Methods to measure RNA expression are the following: Serial Analysis of Gene Expression (SAGE) or Quantitative real-time Polymerase Chain Reaction (qPCR). By using complementary Desoxyribonucleic acid (cDNA) microarray assays important genome-wide information about changes of gene expression in various cell lines can be found out.

As claimed by Montaña et al.¹⁰, the development of NGS techniques implicates vast amount of data which need to be translated. One important question is to find out how the genotype (the genetic expression of a biological attribute) influences the phenotype. By integrating genome editing systems into their research process, investigators

⁸cf.[Zhao et al. 2016]

⁹cf.[Gasperskaja and Kučinskas 2017]

¹⁰cf.[Montaña et al. 2018]

are able to manipulate virtually any gene in a diverse range of cell types and organisms. To give an example, Gasperskaja et al. and Montaña et al. describe Clustered Regularly Interspaced Short Palindromic Repeats Cas-9 (CRISPR-CAS9). In fact, this genome analysis includes generating a direct cut in the double strand of DNA by Cas9 nuclease. Cas9 is driven by a single 20-nucleotide RNA strand which marks the direct breakpoint. After cutting the DNA, the repair machinery of the host cell repairs errors and promotes a modification of the original sequence by a mutation (e.g. insertion, deletion, inversion). But why should be changed the shape of DNA? In the opinion of Montaña et al.¹¹, the use of genetically modified cell lines and animal models help us to better understand the functions of genes and their pathogenesis in diseases, such as cancer.

5.3 Data sources: NCBI and Ensembl genome browser

96

NCBI NCBI was established in Nov. 4, 1988 as a national resource for molecular biology information¹². Its main purpose was to create automated systems for storing and analyzing knowledge about molecular biology as well as biochemistry and genetics. The community of NCBI conducts research on fundamental biomedical problems at a molecular level using mathematical and computational methods. Moreover NCBI collaborates with several National Institutes of Health (NIH) and develops, distributes, supports, and grants access to various databases and software for scientific and medical communities. NCBI also works as an organization for standardization of databases, data deposition and data exchange and biological nomenclature. As a challenge, NCBI needs to find new approaches to deal with the volume and complexity of data. Another complex issue NCBI is facing is how to provide researchers better access to analysis and computing tools. These tools will be useful to understand the genetic legacy and its role in health and disease.

In this article, NCBI was used to query all genomes related to the disease ALL which can be seen in figure 5.1. After selecting the first gene 'IKZF1', the web interface

¹¹cf.[Montaña et al. 2018]

¹²cf.[Biotechnology et al. 2019]

shows details about the given genome sequence (see figure 5.2). Below the given point 'See related', a link to the related source from Ensembl genome browser 96 can be clicked. The next paragraph describes the information from the Ensembl genome browser 96.

NCBI Resources How To Sign in to NCBI

Gene (acute lymphoblastic leukemia) AND "Homo sapiens" Search

Gene sources: Genomic, Mitochondria, Organelles

Categories: Alternatively spliced, Annotated genes, Non-coding, Protein-coding

Sequence content: CCDS, Ensembl, RefSeq, RefSeqGene

Status: Current

Search results: Items: 1 to 20 of 588

See also 3 discontinued or replaced items.

Name/Gene ID	Description	Location	Aliases	MIM
IKZF1 ID: 10320	IKAROS family zinc finger 1 [<i>Homo sapiens</i> (human)]	Chromosome 7, NC_000007.14 (50303453..50405101)	CVID13, Hs.54452, IK1, IKAROS, LYF1, LyF-1, PPP1R92, PRO0758, ZNFN1A1	603023
ARID5B ID: 84159	AT-rich interaction domain 5B [<i>Homo sapiens</i> (human)]	Chromosome 10, NC_000010.11 (61901254..62096948)	DESRT, MRF-2, MRF2	608538
CEBPE ID: 1053	CCAAT enhancer binding protein epsilon [<i>Homo sapiens</i> (human)]	Chromosome 14, NC_000014.9 (23117306..23119611, complement)	C/EBP-epsilon, CRP1	600749
DDC ID: 1644	dopa decarboxylase [<i>Homo sapiens</i> (human)]	Chromosome 7, NC_000007.14 (50458436..50565460, complement)	AADC	107930
TSPAN7 ID: 7102	tetraspanin 7 [<i>Homo sapiens</i> (human)]	Chromosome X, NC_000023.11 (38561478..38688918)	A15, CCG-B7, CD231, DXS1692E, MRX58, MXS1, TALLA-1, TM4SF2, TM4SF2L	300096

Filters: Manage Filters

Results by taxon: Top Organisms [Tree](#)
Homo sapiens (587)
Human mastadenovirus A (1)

Find related data: Database: [Select] Find items

Search details: acute lymphoblastic leukemia[All Fields] AND "Homo sapiens"[All Fields] AND alive[prop] Search See more...

Recent activity: Turn Off Clear

Figure 5.1: NCBI query to find all ALL related gene sequences

IKZF1 IKAROS family zinc finger 1 [*Homo sapiens* (human)]

Gene ID: 10320, updated on 4-Jun-2019

Summary

Official Symbol [IKZF1](#) provided by [HGNC](#)

Official Full Name IKAROS family zinc finger 1 provided by [HGNC](#)

Primary source [HGNC:HGNC:13176](#)

See related [Ensembl:ENSG00000185811](#) [MIM:603023](#)

Gene type protein coding

RefSeq status REVIEWED

Organism [Homo sapiens](#)

Lineage Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo

Also known as IK1; LYF1; LyF-1; CVID13; IKAROS; PPP1R92; PRO0758; ZNFN1A1; Hs.54452

Summary This gene encodes a transcription factor that belongs to the family of zinc-finger DNA-binding proteins associated with chromatin remodeling. The expression of this protein is restricted to the fetal and adult hemo-lymphopoietic system, and it functions as a regulator of lymphocyte differentiation. Several alternatively spliced transcript variants encoding different isoforms have been described for this gene. Most isoforms share a common C-terminal domain, which contains two zinc finger motifs that are required for hetero- or homo-dimerization, and for interactions with other proteins. The isoforms, however, differ in the number of N-terminal zinc finger motifs that bind DNA and in nuclear localization signal presence, resulting in members with and without DNA-binding properties. Only a few isoforms contain the requisite three or more N-terminal zinc motifs that confer high affinity binding to a specific core DNA sequence element in the promoters of target genes. The non-DNA-binding isoforms are largely found in the cytoplasm, and are thought to function as dominant-negative factors. Overexpression of some dominant-negative isoforms have been associated with B-cell malignancies, such as acute lymphoblastic leukemia (ALL). [provided by RefSeq, May 2014]

Expression Biased expression in lymph node (RPKM 13.3), appendix (RPKM 8.6) and 10 other tissues [See more](#)

Orthologs [mouse](#) [all](#)

Table of contents

- Summary
- Genomic context
- Genomic regions, transcripts, and products
- Expression
- Bibliography
- Phenotypes
- Variation
- HIV-1 interactions
- Pathways from BioSystems
- Interactions
- General gene information
 - Markers, Homology, Gene Ontology
- General protein information
- NCBI Reference Sequences (RefSeq)
- Related sequences
- Additional links

Genome Browsers

- Genome Data Viewer
- Variation Viewer (GRCh37.p13)
- Variation Viewer (GRCh38)

Figure 5.2: Detailed information about selected gene sequence

Ensembl genome browser 96 The second genome database used in this article, was the Ensembl genome browser 96¹³. It is a special genome browser for vertebrate genomes. In comparison to NCBI, Ensembl annotates genes and computes multiple alignments and predicts regulatory function. Besides, Ensembl collects disease data and supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. In April 2019, Ensembl published a new release which included a Representational State Transfer (REST) and File Transfer Protocol (FTP) Application Programming Interface (API). Apart from new genomes, Ensembl Release 96 offers a new interface to configure the regulation tracks.

Figure 5.3 shows the Ensembl result of the selected gene from NCBI database.

The screenshot shows the Ensembl genome browser interface for the gene **IKZF1** (ENSG00000185811). The top navigation bar includes links for BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and Blog. A search bar is present on the right. The left sidebar shows a tree of navigation options under 'Human (GRCh38.p12)'. The main content area is divided into sections: 'Gene-based displays' (Summary, Splice variants, Transcript comparison, Gene alleles, Sequence, Secondary Structure, Comparative Genomics, Genomic alignments, Gene tree, Gene gain/loss tree, Orthologues, Paralogs, Ensembl protein families, Ontologies, GO: Biological process, GO: Molecular function, GO: Cellular component, Phenotypes, Genetic Variation, Variant table, Variant image, Structural variants, Gene expression, Pathway, Regulation, External references, Supporting evidence, ID History, Gene history), 'Gene: IKZF1 ENSG00000185811', 'Description' (IKAROS family zinc finger 1 [Source:HGNC Symbol;Acc:HGNC:13176]), 'Gene Synonyms' (Hs.54452, IKAROS, LyF-1, PPP1R92, ZNFN1A1, hlk-1), 'Location' (Chromosome 7: 50,304,068-50,405,101 forward strand. GRCh38:CM000669.2), 'About this gene' (This gene has 23 transcripts (splice variants), 182 orthologues, 11 paralogs, is a member of 1 Ensembl protein family and is associated with 68 phenotypes), 'Transcripts' (Show transcript table), 'Summary' (Name: IKZF1 (HGNC Symbol), CCDS: This gene is a member of the Human CCDS set: CCDS59055.1, CCDS69299.1, CCDS75596.1, CCDS75597.1, CCDS78232.1, CCDS78233.1, CCDS87502.1, UniProtKB: This gene has proteins that correspond to the following UniProtKB identifiers: Q13422, RefSeq: This Ensembl/Genome gene contains transcript(s) for which we have selected identical RefSeq transcript(s). If there are other RefSeq transcripts available they will be in the External references table, Ensembl version: ENSG00000185811.18, Other assemblies: There is no ungapped mapping of this gene onto the GRCh37 assembly. View this locus in the GRCh37 archive: ENSG00000185811, Gene type: Protein coding, Annotation method: Annotation for this gene includes both automatic annotation from Ensembl and Havana manual curation, see).

Figure 5.3: Information about genome IKZF1, found in Ensembl genome browser 96

¹³[Ensembl genome browser 96 2019]

6 Development of a solution for genetic analysis of ALL genomes by implementing LDA

6.1 Problems and challenges of genetic analysis

First of all, importing all the needed data from NCBI was very easy. The NCBI website supports SQL-similar syntaxes so that the query could be as easy as follows: 'acute lymphoblastic leukemia 'AND' HOMO SAPIENS' (to only get the human genomes). By the way, the user of NCBI can choose between many different types of animal genomes.

After importing the data by using pandas, a dataframe object was created and all the relevant data (columns) were filtered.

Furthermore, the most important words that should be clustered or put into topics, where compound expressions, such as 'enhancer binding protein'. One idea was to create bigrams or trigrams within the dictionary and to get all the appearances of these special expressions.

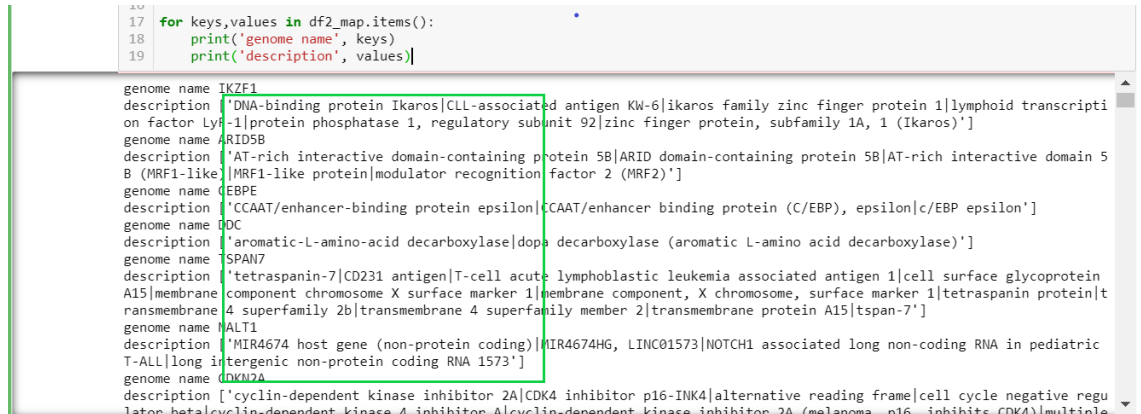
Is the data complete, which includes that it contains all required genomes which can cause ALL. To be more precise, the columns symbol and other designations were very interesting since they gave insights about the core characteristics of each genome. By connecting each genome name to its description, a key value map was created automatically. During the process of data retrieval, there came up the problem of different forms of data and data types. Moreover, there was the 'standard' stopword list from NLTK or spacy which did not include specific words, such as 'DNA-binding protein'. So, these had to be extended by using a custom stopword list. Another problem

that came up was the quality of the data. During the analyzation process, data was extracted from the whole dataset. As can be seen in figure 6.1, every genome was described differently. Some are described by a combination of numbers and letters (e.g. 'tetraspanin-7'), others a described by their functionality (e.g. 'DNA-binding protein Ikaros'). These differences made the analysis and data preparation very challenging, since every description should be compared to the others.

```

17 for keys, values in df2_map.items():
18     print('genome name', keys)
19     print('description', values)

```



The screenshot shows a Jupyter Notebook interface. At the top, there is a code cell with three lines of Python code: a for loop that iterates over items in a dictionary named df2_map, printing the 'genome name' and 'description' for each item. Below the code cell, the output of the code is displayed. It shows several lines of text, each representing a genome and its description. A green rectangle is drawn around a portion of the output text, specifically around the descriptions of 'IKZF1', 'ARID5B', 'EBP', and 'CDK4'. The text within the green rectangle includes descriptions like 'DNA-binding protein Ikaros', 'AT-rich interactive domain-containing protein 5B', 'CCAAT/enhancer binding protein epsilon', and 'aromatic-L-amino-acid decarboxylase'.

Figure 6.1: Extracted data (including genome's name and description), the green rectangle includes the descriptions

6.2 First steps: Draft of developed solution

To get the disease related data, NCBI ¹ was used to get all currently detected mutations of genomes which can cause LDA. The first idea was to build a parser, which iterates over the found 582 genomes. During the iteration, four pre-defined topics are built. The topics included 'the most common nucleotide in the sequence', 'the nucleotide followed by its complementary nucleotide', 'the nucleotide followed by other nucleotides (excepts its complementary)' and 'the double distribution of the same nucleotide' (see figure 6.2).

After the iteration, it compares the oncogenes with the healthy genomes and to figure out where the differences are. The results could be displayed in a diagram. After that, the idea was to create clusters among the two groups or run the LDA algorithm. Due to research, this approach was very naive and supervised because it does not give relevant information about the genomes' mutations. Moreover, there

¹cf.[Biotechnology et al. 2019]

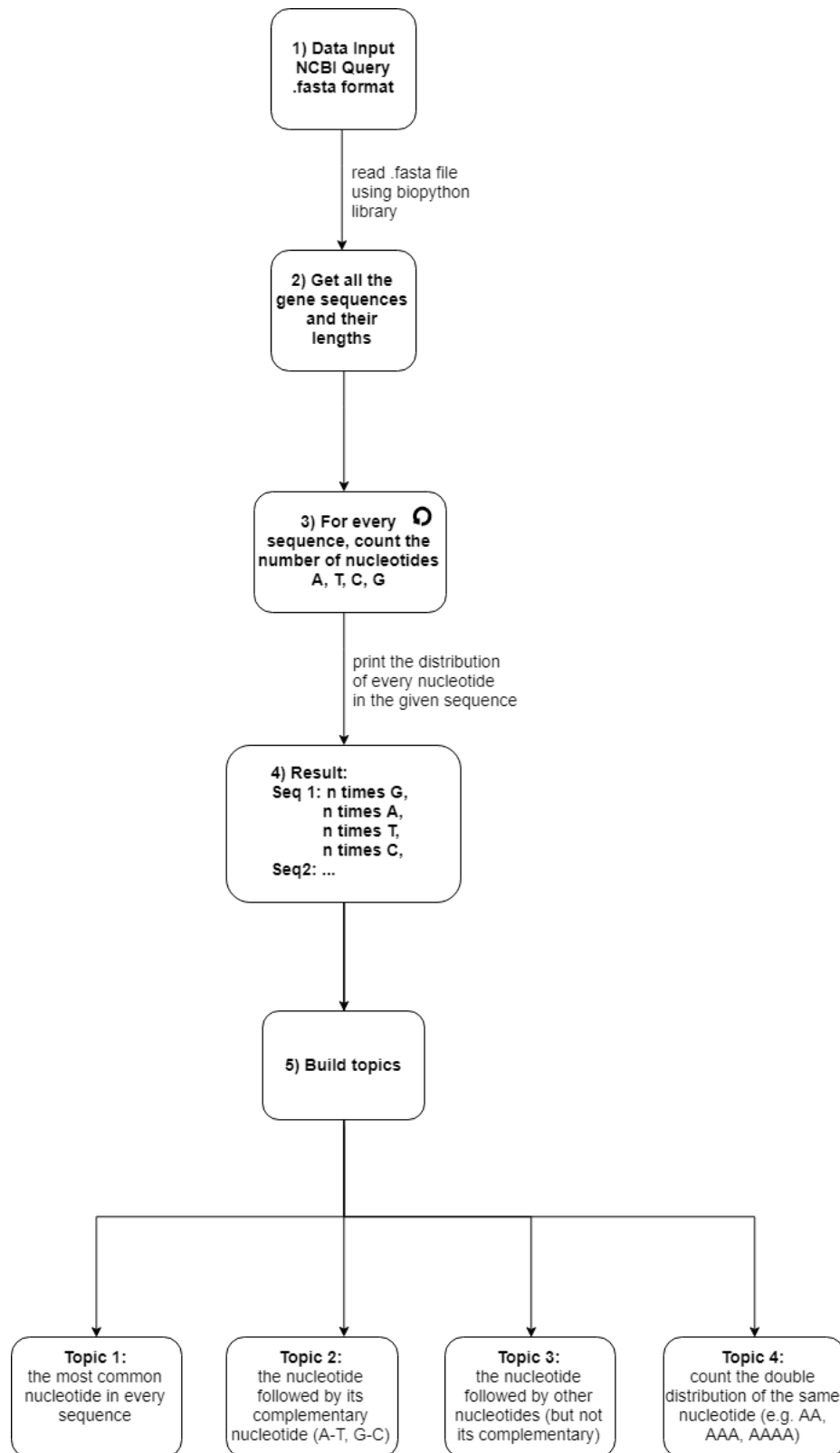


Figure 6.2: Diagram of first draft of developed algorithm

exist many sequencing techniques like Basic Local Alignment Search Tool (BLAST) to parse genome sequences and to find out alignments (similar sequences) between genes. For that reason, the first algorithm 6.2 was not important for the research question.

6.3 Proposed solution

The proposed solution can be divided into two processes: The preprocessing and the central LDA process. In the following both processes are described in detail showing diagrams. Figure 6.3 shows the implemented solution. First of all, the required libraries and python packages have to be imported. After that, the dataset (genome data from NCBI) is converted into a pandas dataframe (table-like data structure). In the next step, only the relevant data is extracted from the dataset. Step five summarizes the LDA process (which is described in detail in figure 6.5). In the last step, the topics are drawn in a topic distribution diagram using the library 'PyLDAvis'. An example of the diagram can be seen in figure 6.4. PyLDAvis gives an overview of all large topics and draws them as large circles. Each circle can be clicked and shows the related words of the selected topic.

After the 'Preprocessing' step, the second process includes several NLP techniques, such as lemmatization, defining stopwords and removing them (see figure 6.5). Since the dataset differentiates from natural language, there had to be added genome-specific stopwords (see step four). With the rest of the data, a dictionary was created. This is relevant for the next step: define the number of topics and call the `LdaModel()` method from the gensim library. There are several parameters, such as the number of topics (which should be adjusted to the size of the dataset) or the number of passes (which means how often the dataset is being parsed) which can be configured. The result of this operation includes the most common words per topic multiplied with their distribution factor (dirichlet factor, see figure 6.6).

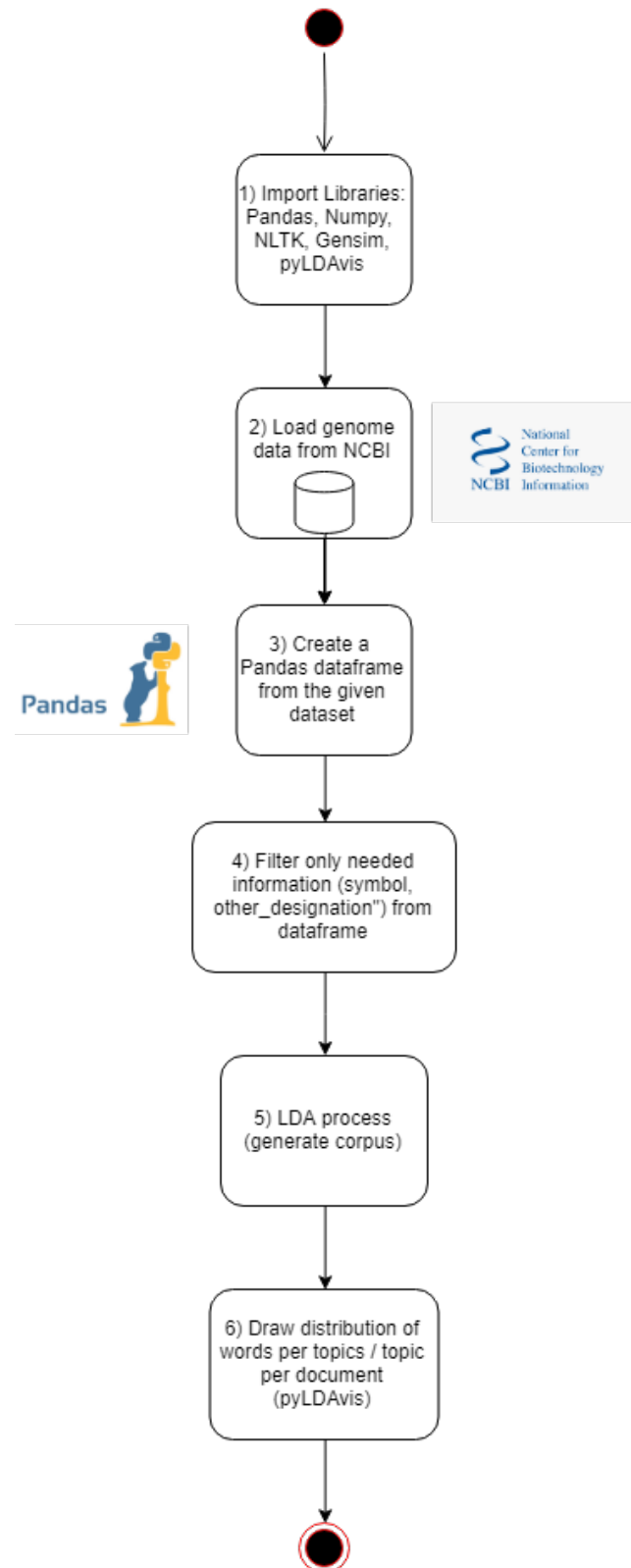


Figure 6.3: Diagram of developed algorithm to create basic topics among given genomes

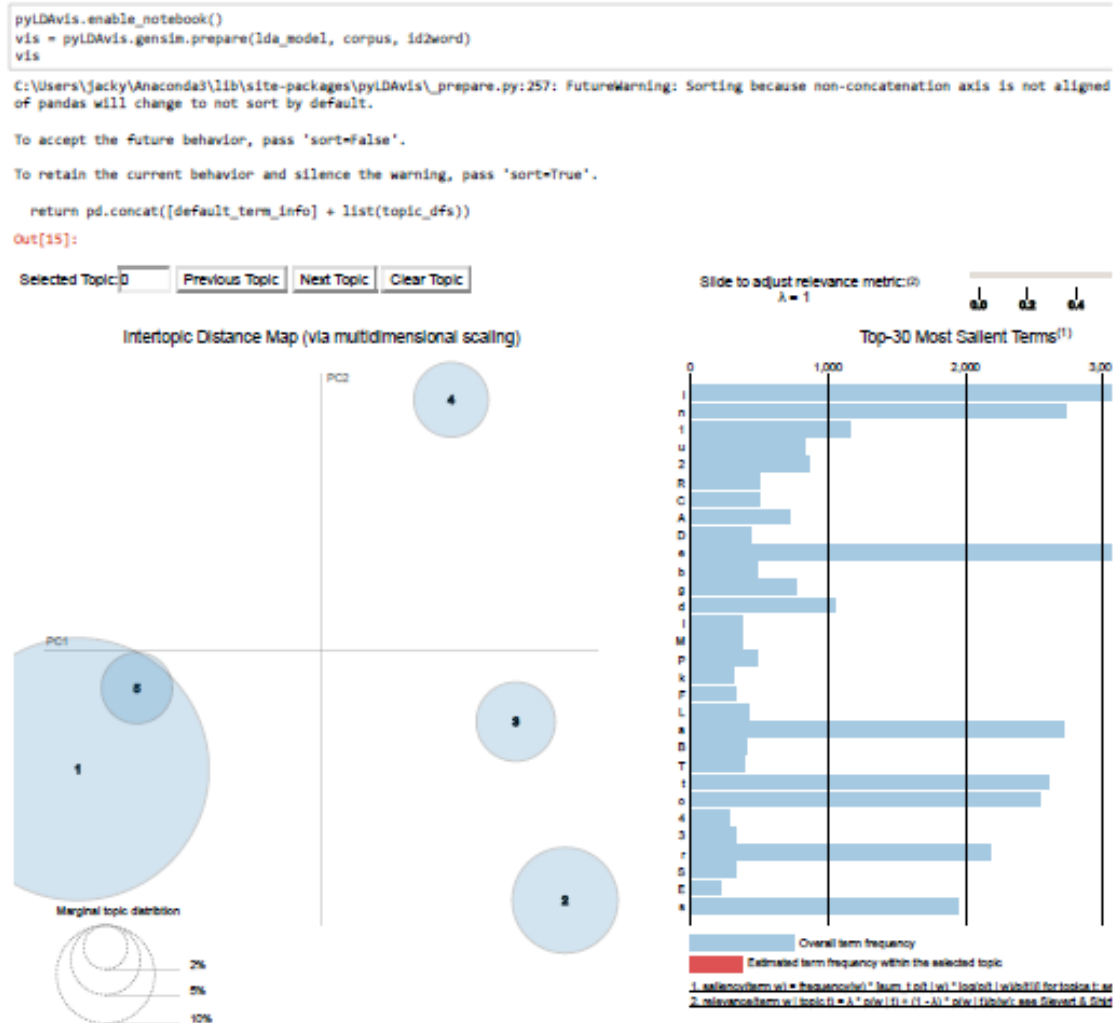


Figure 6.4: Diagram (created with PyLDAvis) which shows the distribution of the topics

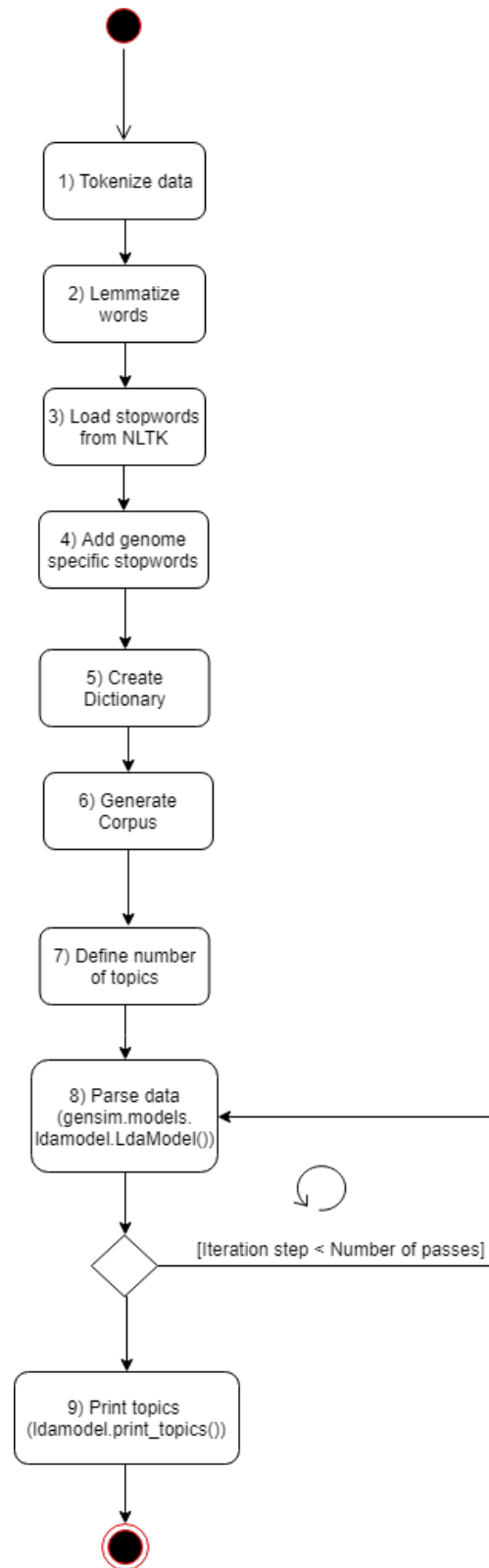


Figure 6.5: Diagram of LDA process

Build LDA Model (using corpus and dictionary)

- 5 topics
- 10 passes

In [12]:

```
# Build LDA model
lda_model = gensim.models.ldamodel.LdaModel(corpus=corpus,
                                             id2word=id2word,
                                             num_topics=5,
                                             random_state=100,
                                             update_every=1,
                                             chunksize=100,
                                             passes=10,
                                             alpha='auto',
                                             per_word_topics=True)
```

In [13]:

```
# Print the Keyword in the 10 topics
print(lda_model.print_topics())
doc_lda = lda_model[corpus]

[(0, '0.248**i" + 0.165**n" + 0.137**k" + 0.097**b" + 0.095**2" + 0.083**u" + 0.055**v" + 0.048**g" + 0.036**d" + 0.029**q"', (1, '0.226**l" + 0.076**i" + 0.065**3" + 0.063**5" + 0.048**h" + 0.044**5" + 0.038**0"', (2, '0.117**e" + 0.091**i" + 0.086**a" + 0.082**t" + 0.082**l" + 0.061**s" + 0.052**c"', (3, '0.177**c" + 0.155**d" + 0.148**2" + 0.098**4" + 0.098**8" + 0.069**9" + 0.068**A" + 0.064**K" + 0.0 + 0.148**I" + 0.146**M" + 0.128**F" + 0.085**E" + 0.085**2" + 0.084**6" + 0.065**7" + 0.029**H" + 0.019**Y"')])]
```

Figure 6.6: Call of method LdaModel() from library Gensim and its output

6.4 Results

Other than expected, the results of the genome analysis by implenting LDA topic modelling, gave no meaningful information about the similarities between the given genomes. The topics were not very precise and the words per topic were very distinctive (nouns were mixed with verbs and numbers). Moreover, as stated in section 6.1, the genomes were described more or less detailed which led to a heterogeneous dataset. This leads to the problem, that the analyzed and processed data cannot be interpreted neither in a biological nor in a data-scientific way. If the results were more precise, it could have been possible to state that for example the majority of the genomes (which are words of the largest topic) are 'suppressor' genomes which inhibit the functionality of enzymes. But unfortunately, there cannot be made any logical statement about the results of analyzation.

7 Conclusion and Outlook

7.1 Lessons learned

There are many genome analysis tools in the internet. NIH provides several interfaces for genome alignment methods to find out the similarity between two given genomes ¹. But in the given case, all given gene sequences should be compared with each other in order to find appropriate clusters. Another point is the algorithm of LDA. Since it is an unsupervised algorithm and is more convenient for natural language than for genome-specific data, it is difficult to build topics among the genome specific data and measure their similarities.

7.2 Conclusion

As a conclusion, LDA gives the possibility to analyse the natural language by creating topics among a given dataset. Moreover, it creates a probability distribution of words per topic. The found topics can be summarized in a document and can also be stated as a distribution of topics per document. In this paper, there was only 'one' document, a dataset of genomes, which should be clustered automatically into topics. During the development and implementation of the LDA algorithm, there came up many problems and challenges (see section 6.1), such as to extend the stopwords list. The results of the LDA process were very unprecise and difficult to interpret. This might be caused by the algorithm itself. Research revealed that its more appropriate to natural language analysis than to genome specific data. Furthermore, another assumption is that the LDA process is too costly for a general analysis of genome data.

¹cf.[NCBI BLAST ; Sequence Similarity Searching ; EMBL-EBI 2019]

7.3 Outlook

As described in 7.1, the problem of unsupervised clustering algorithm (which is more suitable for natural language processing) and the genome-sepcific data came up. As a proposal for future analysis, a supervised technique like K-Means clustering could be used to better identify similar genomes and to build topics.

Another solution can be to prepare a neural network which is fed with genome test data and trained for some iterations. After that, the LDA process is executed on test data and creates more precise topics.

After that, to propose an easier analysis method, the similarity between the genomes could be calculated using certain 'distance metrics', such as 'sklearn.neighbors.DistanceMetric'. In order to use this method, the words have to be converted into vectors, using the NLTK method 'word2vec()'. Consequently, running the sklearn command should create useful results which can be interpreted better than the results of LDA.

8 Abbreviations

LDA	Latent Dirichlet Allocation
NCBI	National Center for Biotechnology Information
ALL	Acute Lymphoblastic Leukemia
AML	Acute Myeloid Leukemia
CLL	Chronic Lymphocytic Leukemia
CML	Chronic Myeloid Leukemia
MSA	Multiple Sequence Alignment
MUSCLE	Multiple Sequence Comparison by Log- Expectation
DNA	Desoxyribonucleic acid
NGS	Next-Generation Sequencing
CDNA	complementary Desoxyribonucleic acid
IR	Information Retrieval
NER	named entity recognition
IE	Information extraction
KD	Knowledge Discovery
NLP	Natural Language Processing
RNA	Ribonucleic Acid
CRNA	complementary Ribonucleic Acid

CART Classification and Regression Tree

ID3 Iterative Dichotomized 3

DT Decision Tree

PCA Principal Component Analysis

RST Rhethorical Structure Theory

HMM Hidden-Markov-Models

BM Bayesian models

RE Relation extraction

MIRNA micro Ribonucleic Acid

CNV copy number variant

NLTK Natural Language Toolkit

CRISPR-CAS9 Clustered Regularly Interspaced Short Palindromic Repeats Cas-9

SNP Single Nucleotide Polymorphism

MRNA messenger RNA

PCR Polymerase Chain Reaction

SAGE Serial Analysis of Gene Expression

QPCR Quantitative real-time Polymerase Chain Reaction

CAR Chimeric Antigen Receptors

MCMC Markov Chain Monte Carlo

ARM Association Rule Mining

MCLDA Multiple-channel Latent Dirichlet Allocation

CGS Collapsed Gibbs model

TFIDF term frequency inverse document frequency

NIH	National Institutes of Health
REST	Representational State Transfer
FTP	File Transfer Protocol
API	Application Programming Interface
CALLCG	Childhood Acute Lymphoblastic Leukemia Collaborative Group
EFS	event-free survival
T-ALL	T-Acute Lymphoblastic Leukemia
BCP-ALL	B-cell precursor Acute Lymphoblastic Leukemia
CNA	DNA copy-number alterations
BLAST	Basic Local Alignment Search Tool

Bibliography

- Andersson, Anna et al. (Mar. 2016). "Acute lymphoblastic leukaemia". en. In: *The Genetic Basis of Haematological Cancers*. Ed. by Sabrina Tosi and Alistair G. Reid. Chichester, UK: John Wiley & Sons, Ltd, pp. 223–264. ISBN: 978-1-118-52794-8. DOI: 10.1002/9781118527948.ch5. URL: <http://doi.wiley.com/10.1002/9781118527948.ch5> (visited on 05/31/2019).
- Biotechnology, National Center for et al. (2019). *National Center for Biotechnology Information*. en. URL: <https://www.ncbi.nlm.nih.gov/> (visited on 05/09/2019).
- Blei, David M., Andrew Y. Ng, and Michael I. Jordan (2003). "Latent Dirichlet Allocation". In: *Journal of Machine Learning Research* 3.Jan, pp. 993–1022. ISSN: ISSN 1533-7928. URL: <http://www.jmlr.org/papers/v3/blei03a.html> (visited on 04/29/2019).
- Djatna, Taufik, Medria Kusuma Dewi Hardhienata, and Anis Fitri Nur Masruriyah (Dec. 2018). "An intuitionistic fuzzy diagnosis analytics for stroke disease". en. In: *Journal of Big Data* 5.1, p. 35. ISSN: 2196-1115. DOI: 10.1186/s40537-018-0142-7. URL: <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-018-0142-7> (visited on 04/07/2019).
- Ensembl genome browser 96 (2019). URL: <https://www.ensembl.org/index.html> (visited on 06/02/2019).
- Gasperskaja, Evelina and Vaidutis Kučinskas (2017). "The most common technologies and tools for functional genome analysis". In: *Acta Medica Lituanica* 24.1, pp. 1–11. ISSN: 1392-0138. DOI: 10.6001/actamedica.v24i1.3457. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5467957/> (visited on 05/09/2019).
- Group (CALLCG), Childhood Acute Lymphoblastic Leukaemia Collaborative (2009). "Beneficial and harmful effects of anthracyclines in the treatment of childhood acute lymphoblastic leukaemia: a systematic review and meta-analysis". en. In: *British Journal of Haematology* 145.3, pp. 376–388. ISSN: 1365-2141. DOI: 10.1111/j.

- 1365–2141.2009.07624.x. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2141.2009.07624.x> (visited on 05/31/2019).
- Hoffman, Matthew, Francis R. Bach, and David M. Blei (2010). “Online Learning for Latent Dirichlet Allocation”. In: *Advances in Neural Information Processing Systems* 23. Ed. by J. D. Lafferty et al. Curran Associates, Inc., pp. 856–864. URL: <http://papers.nips.cc/paper/3902-online-learning-for-latent-dirichlet-allocation.pdf> (visited on 04/19/2019).
- Jagadev, P. and H. G. Virani (May 2017). “Detection of leukemia and its types using image processing and machine learning”. In: *2017 International Conference on Trends in Electronics and Informatics (ICEI)*, pp. 522–526. doi: 10.1109/ICEI.2017.8300983.
- Jurca, Gabriela et al. (Dec. 2016). “Integrating text mining, data mining, and network analysis for identifying genetic breast cancer trends”. en. In: *BMC Research Notes* 9.1. ISSN: 1756-0500. doi: 10.1186/s13104-016-2023-5. URL: <http://bmresnotes.biomedcentral.com/articles/10.1186/s13104-016-2023-5> (visited on 04/06/2019).
- Lee, Hyunju, Sek Won Kong, and Peter J. Park (Apr. 2008). “Integrative analysis reveals the direct and indirect interactions between DNA copy number aberrations and gene expression changes”. eng. In: *Bioinformatics (Oxford, England)* 24.7, pp. 889–896. ISSN: 1367-4811. doi: 10.1093/bioinformatics/btn034.
- Lloret, Elena and Manuel Palomar (Jan. 2012). “Text summarisation in progress: a literature review”. In: *Artificial Intelligence Review* 37.1, pp. 1–41. ISSN: 1573-7462. doi: 10.1007/s10462-011-9216-z. URL: <https://doi.org/10.1007/s10462-011-9216-z>.
- Lu, Hsin-Min, Chih-Ping Wei, and Fei-Yuan Hsiao (Apr. 2016). “Modeling healthcare data using multiple-channel latent Dirichlet allocation”. In: *Journal of Biomedical Informatics* 60, pp. 210–223. ISSN: 1532-0464. doi: 10.1016/j.jbi.2016.02.003. URL: <http://www.sciencedirect.com/science/article/pii/S1532046416000253> (visited on 04/29/2019).
- Montaño, Adrián et al. (July 2018). “Targeted genome editing in acute lymphoblastic leukemia: a review”. In: *BMC Biotechnology* 18.1, p. 45. ISSN: 1472-6750. doi: 10.1186/s12896-018-0455-9. URL: <https://doi.org/10.1186/s12896-018-0455-9> (visited on 04/19/2019).

- Natural Language Toolkit — NLTK 3.4.1 documentation* (2019). URL: <https://www.nltk.org/> (visited on 05/09/2019).
- NCBI BLAST ; Sequence Similarity Searching ; EMBL-EBI* (2019). URL: <https://www.ebi.ac.uk/Tools/sss/ncbiblast/nucleotide.html> (visited on 06/06/2019).
- Park, Laurence A. F. and Kotagiri Ramamohanarao (2009). "The Sensitivity of Latent Dirichlet Allocation for Information Retrieval". en. In: *Machine Learning and Knowledge Discovery in Databases*. Ed. by Wray Buntine et al. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 176–188. ISBN: 978-3-642-04174-7.
- Řehůřek, Radim and Petr Sojka (May 2010). "Software Framework for Topic Modelling with Large Corpora". English. In: *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. <http://is.muni.cz/publication/884893/en>. Valletta, Malta: ELRA, pp. 45–50.
- "Systematic review of the addition of vincristine plus steroid pulses in maintenance treatment for childhood acute lymphoblastic leukaemia – an individual patient data meta-analysis involving 5659 children" (2010). en. In: *British Journal of Haematology* 149.5, pp. 722–733. ISSN: 1365-2141. DOI: 10.1111/j.1365-2141.2010.08148.x. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2141.2010.08148.x> (visited on 05/31/2019).
- Tang, W., H. Cao, and Y. Wang (July 2011). "Subtyping of Leukemia with Gene Expression Analysis Using Compressive Sensing Method". In: *2011 IEEE First International Conference on Healthcare Informatics, Imaging and Systems Biology*, pp. 76–80. DOI: 10.1109/HISB.2011.60.
- Twinandilla, Shiva et al. (Jan. 2018). "Multi-Document Summarization Using K-Means and Latent Dirichlet Allocation (LDA) – Significance Sentences". In: *Procedia Computer Science*. The 3rd International Conference on Computer Science and Computational Intelligence (ICCSCI 2018) : Empowering Smart Technology in Digital Era for a Better Life 135, pp. 663–670. ISSN: 1877-0509. DOI: 10.1016/j.procs.2018.08.220. URL: <http://www.sciencedirect.com/science/article/pii/S1877050918315138> (visited on 04/29/2019).
- Wang, Xiaogang and Eric Grimson (2008). "Spatial Latent Dirichlet Allocation". In: *Advances in Neural Information Processing Systems 20*. Ed. by J. C. Platt et al. Curran Associates, Inc., pp. 1577–1584. URL: <http://papers.nips.cc/paper/3278-spatial-latent-dirichlet-allocation.pdf> (visited on 04/19/2019).

- Xu, Yahong et al. (Sept. 2013). "RETRACTED ARTICLE: Integrated gene network analysis and text mining revealing PIK3R1 regulated by miR-127 in human bladder cancer". In: *European Journal of Medical Research* 18.1, p. 29. ISSN: 2047-783X. DOI: 10.1186/2047-783X-18-29. URL: <https://doi.org/10.1186/2047-783X-18-29>.
- Yang, Xi et al. (Dec. 2018). "Constructing a database for the relations between CNV and human genetic diseases via systematic text mining". en. In: *BMC Bioinformatics* 19.S19, p. 528. ISSN: 1471-2105. DOI: 10.1186/s12859-018-2526-2. URL: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-018-2526-2> (visited on 04/11/2019).
- Zhao, Weizhong et al. (May 2016). "A novel procedure on next generation sequencing data analysis using text mining algorithm". In: *BMC Bioinformatics* 17.1, p. 213. ISSN: 1471-2105. DOI: 10.1186/s12859-016-1075-9. URL: <https://doi.org/10.1186/s12859-016-1075-9>.

9 Appendix A

Ehrenwörtliche Erklärung

Hiermit versichere ich, dass die vorliegende Arbeit von mir selbstständig und ohne unerlaubte Hilfe angefertigt worden ist, insbesondere dass ich alle Stellen, die wörtlich oder annähernd wörtlich aus Veröffentlichungen entnommen sind, durch Zitate als solche gekennzeichnet habe. Ich versichere auch, dass die von mir eingereichte schriftliche Version mit der digitalen Version übereinstimmt. Weiterhin erkläre ich, dass die Arbeit in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde / Prüfungsstelle vorgelegen hat. Ich erkläre mich damit nicht einverstanden, dass die Arbeit der Öffentlichkeit zugänglich gemacht wird. Ich erkläre mich damit einverstanden, dass die Digitalversion dieser Arbeit zwecks Plagiatsprüfung auf die Server externer Anbieter hochgeladen werden darf. Die Plagiatsprüfung stellt keine Zurverfügungstellung für die Öffentlichkeit dar.

Ort, Datum (Vorname Nachname)

10 Appendix B

Titel der Arbeit: Development of a solution for genetic analysis of ALL genomes by implementing Latent Dirichlet Allocation

Name: Jacqueline Franßen

Matrikelnummer: 496804

Data Preparation ¶

- Imports (Pandas, Numpy, Spacy, NLTK, Gensim)
- Create a dataframe object from given dataset
- Filter Dataframe for relevant data for LDA

In [1]:

```
import pandas as pd
import numpy as np
import re
import spacy

import nltk
from nltk.corpus import stopwords
stop_words = stopwords.words('english')
nltk.download('wordnet')
from nltk.corpus import stopwords
nltk.download('stopwords')
from nltk.tokenize import word_tokenize
nltk.download('punkt')
from nltk import word_tokenize, sent_tokenize

from nltk.corpus import wordnet as wn
from nltk.stem.wordnet import WordNetLemmatizer
import random
from gensim import corpora
import pickle
import gensim
from gensim.models import CoherenceModel
from gensim.test.utils import common_texts
from gensim.corpora.dictionary import Dictionary
import pyLDAvis.gensim

[nltk_data] Downloading package wordnet to
[nltk_data]   C:\Users\jacky\AppData\Roaming\nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data]   C:\Users\jacky\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package punkt to
[nltk_data]   C:\Users\jacky\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
```

Data import

- Create a dataframe object from given dataset

In [2]:

```
df1 = pd.read_csv('all_genes.txt', error_bad_lines=False, sep='\t', comment='#', low_memory=False, header=0)
df1.head()
df1
```

Out[2]:

	tax_id	Org_name	GeneID	CurrentID	Status	Symbol	Aliases	description	other_designations	map_location	chromosome	ge
0	9606	Homo sapiens	10320	0	live	IKZF1	CVID13, Hs.54452, IK1, IKAROS, LYF1, LyF-1, PP...	IKAROS family zinc finger 1	DNA-binding protein Ikaros CLL-associated anti...	7p12.2		7
1	9606	Homo sapiens	84159	0	live	ARID5B	DESRT, MRF-2, MRF2	AT-rich interaction domain 5B	AT-rich interactive domain-containing protein ...	10q21.2		10
2	9606	Homo sapiens	1053	0	live	CEBPE	C/EBP-epsilon, CRP1	CCAAT enhancer binding protein epsilon	CCAAT/enhancer-binding protein epsilon CCAAT/e...	14q11.2		14
3	9606	Homo sapiens	1644	0	live	DDC	AADC	dopa decarboxylase	aromatic-L-amino-acid decarboxylase dopa decar...	7p12.2-p12.1		7
4	9606	Homo sapiens	7102	0	live	TSPAN7	A15, CCG-B7, CD231, DXS1692E, MRX58, MXS1, TAL...	tetraspanin 7	tetraspanin-7 CD231 antigen T-cell acute lymph...	Xp11.4		X
5	9606	Homo sapiens	8009	0	live	LALL	NaN	Lymphomatous acute lymphoblastic leukemia	NaN	9p22-p21		9
6	9606	Homo sapiens	101928483	0	live	NALT1	LINC01573, MIR4674HG, NALT, TCONS_id_00029132	NOTCH1 associated lncRNA in T cell acute lymph...	MIR4674 host gene (non-protein coding) MIR4674...	9q34.3		9
7	9606	Homo sapiens	338436	0	live	BLACE	NaN	B cell acute lymphoblastic leukemia expressed	NaN	7q36.3		7
8	9606	Homo sapiens	1029	0	live	CDKN2A	ARF, CDK4I, CDKN2, CMM2, INK4, INK4A, MLM, MTS...	cyclin dependent kinase inhibitor 2A	cyclin-dependent kinase inhibitor 2A CDK4 inhi...	9p21.3		9
9	9606	Homo sapiens	5142	0	live	PDE4B	DPDE4, PDEIVB	phosphodiesterase 4B	cAMP-specific 3',5'-cyclic phosphodiesterase 4...	1p31.3		1
10	9606	Homo sapiens	3600	0	live	IL15	IL-15	interleukin 15	interleukin-15	4q31.21		4
11	9606	Homo sapiens	8626	0	live	TP63	AIS, B(p51A), B(p51B), EEC3, KET, LMS, NBP, OF...	tumor protein p63	tumor protein 63 amplified in squamous cell ca...	3q28		3
12	9606	Homo sapiens	2625	0	live	GATA3	HDR, HDRS	GATA binding protein 3	trans-acting T-cell-specific transcription fac...	10p14		10
13	9606	Homo sapiens	8202	0	live	NCOA3	ACTR, AIB-1, AIB1, CAGH16, CTG26, KAT13B, RAC3...	nuclear receptor coactivator 3	nuclear receptor coactivator 3 CBP-interacting...	20q13.12		20
14	9606	Homo sapiens	6262	0	live	RYR2	ARVC2, ARVD2, RYR-2, RyR, VTSIP	ryanodine receptor 2	ryanodine receptor 2 cardiac muscle ryanodine ...	1q43		1
15	9606	Homo sapiens	3738	0	live	KCNA3	HGK5, HLK3, HPCN3, HUKIII, KV1.3, MK3, PCN3	potassium voltage-gated channel subfamily A me...	potassium voltage-gated channel subfamily A me...	1p13.3		1
16	9606	Homo sapiens	5305	0	live	PIP4K2A	PI5P4KA, PIP5K2A, PIP5KII-alpha, PIP5KIIA, PIPK	phosphatidylinositol-5-phosphate 4-kinase type...	phosphatidylinositol 5-phosphate 4-kinase type...	10p12.2		10
17	9606	Homo sapiens	84441	0	live	MAML2	MAM-3, MAM2, MAM3, MLL-MAML2	mastermind like transcriptional coactivator 2	mastermind-like protein 2 mam-2 mastermind-like 2	11q21		11
18	9606	Homo sapiens	5795	0	live	PTPRJ	CD148, DEP1, HPTPeta, R-PTP-ETA, SCC1	protein tyrosine phosphatase receptor type J	receptor-type tyrosine-protein phosphatase eta...	11p11.2		11
19	9606	Homo sapiens	56288	0	live	PARD3	ASIP, Baz, PAR3, PAR3alpha, PARD-3A, PPP1R118,...	par-3 family cell polarity regulator	partitioning defective 3 homolog CTCL tumor an...	10p11.22-p11.21		10
20	9606	Homo sapiens	59350	0	live	RXFP1	LGR7, RXFP1	relaxin family peptide receptor 1	relaxin receptor 1 leucine-rich repeat-contain...	4q32.1		4
21	9606	Homo sapiens	6916	0	live	TBXAS1	BDPLT14, CYP5, CYP5A1, GHOSAL, THAS, TS, TXAS,...	thromboxane A synthase 1	thromboxane-A synthase TXA synthase cytochrome...	7q34		7
22	9606	Homo sapiens	2887	0	live	GRB10	GRB-IR, Grb-10, IRBP, MEG1, RSS	growth factor receptor bound protein 10	growth factor receptor-bound protein 10 GRB10 ...	7p12.1		7
23	9606	Homo sapiens	9844	0	live	ELMO1	CED-12, CED12, ELMO-1	engulfment and cell motility 1	engulfment and cell motility protein 1 ced-12 ...	7p14.2-p14.1		7
24	9606	Homo sapiens	7157	0	live	TP53	BCC7, BMFS5, LFS1, P53, TRP53	tumor protein p53	cellular tumor antigen p53 antigen NY-CO-13 mu...	17p13.1		17
25	9606	Homo sapiens	9863	0	live	MAGI2	ACVRIP1, AIP-1, AIP1, ARIP1, MAGI-2, NPHS15, S...	membrane associated guanylate kinase, WW and P...	membrane-associated guanylate kinase, WW and P...	7q21.11		7
26	9606	Homo sapiens	2952	0	live	GSTT1	NaN	glutathione S-transferase theta 1	glutathione S-transferase theta-1 GST class-th...	22q11.23		22

	tax_id	Org_name	GeneID	CurrentID	Status	Symbol	Aliases	description	other_designations	map_location	chromosome	ge
27	9606	Homo sapiens	6014	0	live	RIT2	RIBA, RIN, ROC2	Ras like without CAAX 2	GTP-binding protein Rit2 GTP-binding protein R...	18q12.3	18	
28	9606	Homo sapiens	340419	0	live	RSPO2	CRISTIN2, HHRRD, TETAMS2	R-spondin 2	R-spondin-2 R-spondin 2 homolog roof plate-spe...	8q23.1	8	
29	9606	Homo sapiens	5836	0	live	PYGL	GSD6	glycogen phosphorylase L	glycogen phosphorylase, liver form phosphoryla...	14q22.1	14	
...
562	9606	Homo sapiens	127933	0	live	UHMK1	KIS, KIST, P-CIP2	U2AF homology motif kinase 1	serine/threonine-protein kinase Kist KIS prote...	1q23.3	1	
563	9606	Homo sapiens	388585	0	live	HES5	bHLHb38	hes family bHLH transcription factor 5	transcription factor HES-5 class B basic helix...	1p36.32	1	
564	9606	Homo sapiens	55729	0	live	ATF7IP	AM, ATF-IP, MCAF, MCAF1, p621	activating transcription factor 7 interacting ...	activating transcription factor 7-interacting ...	12p13.1	12	
565	9606	Homo sapiens	574504	0	live	MIR502	MIRN502, hsa-mir-502, mir-502	microRNA 502	NaN	Xp11.23	X	
566	9606	Homo sapiens	554212	0	live	MIR448	MIRN448, hsa-mir-448, miRNA448	microRNA 448	NaN	Xq23	X	
567	9606	Homo sapiens	83937	0	live	RASSF4	AD037	Ras association domain family member 4	ras association domain-containing protein 4 Ra...	10q11.21	10	
568	9606	Homo sapiens	414899	0	live	BLID	BRCC2	BH3-like motif containing, cell death inducer	BH3-like motif-containing cell death inducer b...	11q24.1	11	
569	9606	Homo sapiens	79370	0	live	BCL2L14	BCLG	BCL2 like 14	apoptosis facilitator Bcl-2-like protein 14 BC...	12p13.2	12	
570	9606	Homo sapiens	79368	0	live	FCRL2	CD307b, FCRH2, IFGP4, IRTA4, SPAP1, SPAP1A, SP...	Fc receptor like 2	Fc receptor-like protein 2 IFGP family protein...	1q23.1	1	
571	9606	Homo sapiens	266977	0	live	ADGRF1	GPR110, KPG_012, PGR19, hGPCR36	adhesion G protein-coupled receptor F1	adhesion G-protein coupled receptor F1 G prote...	6	6	
572	9606	Homo sapiens	92912	0	live	UBE2Q2	NaN	ubiquitin conjugating enzyme E2 Q2	ubiquitin-conjugating enzyme E2 Q2 E2 ubiquiti...	15q24.2	15	
573	9606	Homo sapiens	115350	0	live	FCRL1	CD307a, FCRH1, IFGP1, IRTA5	Fc receptor like 1	Fc receptor-like protein 1 IFGP family protein...	1q23.1	1	
574	9606	Homo sapiens	80206	0	live	FHOD3	FHOS2, Formactin2	formin homology 2 domain containing 3	FH1/FH2 domain-containing protein 3 formactin-...	18q12.2	18	
575	9606	Homo sapiens	54970	0	live	TTC12	TPARM	tetratricopeptide repeat domain 12	tetratricopeptide repeat protein 12 TPR repeat...	11q23.2	11	
576	9606	Homo sapiens	6887	0	live	TAL2	NaN	TAL bHLH transcription factor 2	T-cell acute lymphocytic leukemia protein 2 T-...	9q31.2	9	
577	9606	Homo sapiens	693197	0	live	MIR612	MIRN612, hsa-mir-612	microRNA 612	NaN	11q13.1	11	
578	9606	Homo sapiens	1389	0	live	CREBL2	NaN	cAMP responsive element binding protein like 2	cAMP-responsive element-binding protein-like 2...	12p13.1	12	
579	9606	Homo sapiens	160365	0	live	CLECL1	DCAL-1, DCAL1	C-type lectin like 1	C-type lectin-like domain family 1 DC-associat...	12p13.31	12	
580	9606	Homo sapiens	128710	0	live	SLX4IP	C20orf94, bA204H22.1, bA254M13.1, dJ1099D15.3	SLX4 interacting protein	protein SLX4IP	20p12.2	20	
581	9606	Homo sapiens	57824	0	live	HMHB1	HB-1, HB-1Y, HLA-HB1	histocompatibility minor HB-1	minor histocompatibility protein HB-1 minor hi...	5q31.3	5	
582	9606	Homo sapiens	162979	0	live	ZNF296	ZFP296, ZNF342	zinc finger protein 296	zinc finger protein 296 zinc finger protein 342	19q13.32	19	
583	9606	Homo sapiens	574493	0	live	MIR520H	MIRN520H	microRNA 520h	hsa-mir-520h	19q13.42	19	
584	9606	Homo sapiens	92241	0	live	RCSD1	CAPZIP, MK2S4	RCSD domain containing 1	capZ-interacting protein RCSD domain-containin...	1q24.2	1	
585	9606	Homo sapiens	107980440	0	live	LOC107980440	NaN	ABL breakpoint recombination region	ABL proto-oncogene 1, non-receptor tyrosine ki...	9q34.1	9	
586	9606	Homo sapiens	107963955	0	live	LOC107963955	NaN	BCR-ABL major-breakpoint cluster region	BCR p210 Philadelphia chromosome recombination...	22q11.23	22	

3.8.2019

lda_all_genomes_v04

	tax_id	Org_name	GeneID	CurrentID	Status	Symbol	Aliases	description	other_designations	map_location	chromosome	ge
587	9606	Homo sapiens	107963951	0	live	LOC107963951	NaN	BCR-ABL minor-breakpoint cluster region	BCR p190 Philadelphia chromosome recombination...	22q11.23	22	
588	9606	Homo sapiens	192343	0	live	NEWENTRY	NaN	Record to support submission of GeneRIFs for a...	NaN	NaN	NaN	
589	9606	Homo sapiens	107648866	0	live	LOC107648866	NaN	meiotic recombination hotspot DNA3	NaN	6p21.3	6	
590	9606	Homo sapiens	3197	0	live	HOXA@	HOX1@	homeobox A cluster	homeo box A cluster	7p15.2	7	
591	129875	Human mastadenovirus A	1460850	0	live	E2A	HAdVAgp12	single-stranded DNA-binding protein	single-stranded DNA-binding protein	NaN	NaN	

592 rows × 18 columns

Only get relevant information from dataframe (2 colummns: Symbol and other_designations)

- Filter Dataframe for relevant data for LDA

In [3]:

```
df2 = pd.read_csv('all_genes.txt',
                  sep='\t', comment='#', low_memory=False,
                  usecols = ["other_designations", "Symbol"],
                  header = 0 )
df2 = df2.dropna() # remove NaN values
# concat the 2 columns to one
df2 = df2['Symbol'].astype(str) + ' ' + df2['other_designations']
df2
```

Out[3]:

```
0      IKZF1 DNA-binding protein Ikaros|CLL-associate...
1      ARID5B AT-rich interactive domain-containing p...
2      CEBPE CCAAT/enhancer-binding protein epsilon|C...
3      DDC aromatic-L-amino-acid decarboxylase|dopa d...
4      TSPAN7 tetraspanin-7|CD231 antigen|T-cell acut...
6      NALT1 MIR4674 host gene (non-protein coding)|M...
8      CDKN2A cyclin-dependent kinase inhibitor 2A|CD...
9      PDE4B cAMP-specific 3',5'-cyclic phosphodieste...
10     IL15 interleukin-15
11     TP63 tumor protein 63|amplified in squamous ce...
12     GATA3 trans-acting T-cell-specific transcripti...
13     NCOA3 nuclear receptor coactivator 3|CBP-inter...
14     RYR2 ryanodine receptor 2|cardiac muscle ryano...
15     KCNA3 potassium voltage-gated channel subfamil...
16     PIP4K2A phosphatidylinositol 5-phosphate 4-kin...
17     MAML2 mastermind-like protein 2|mam-2|mastermi...
18     PTPRJ receptor-type tyrosine-protein phosphata...
19     PARD3 partitioning defective 3 homolog|CTCL tu...
20     RXFP1 relaxin receptor 1|leucine-rich repeat-c...
21     TBXAS1 thromboxane-A synthase|TXA synthase|cyt...
22     GRB10 growth factor receptor-bound protein 10|...
23     ELM01 engulfment and cell motility protein 1|c...
24     TP53 cellular tumor antigen p53|antigen NY-CO...
25     MAGI2 membrane-associated guanylate kinase, WW...
26     GSTT1 glutathione S-transferase theta-1|GST cl...
27     RIT2 GTP-binding protein Rit2|GTP-binding prot...
28     RSP02 R-spondin-2|R-spondin 2 homolog|roof pla...
29     PYGL glycogen phosphorylase, liver form|phosph...
30     MTHFR methylenetetrahydrofolate reductase|5,10...
31     KCNE4 potassium voltage-gated channel subfamil...
...
557    SIAE sialate O-acetyltransferase|H-Lse|cytosolic ...
558    MLXIP MLX-interacting protein|MLX interactor|c...
559    TSGA10 testis-specific gene 10 protein|cancer/...
560    DPH1 2-(3-amino-3-carboxypropyl)histidine synt...
561    TPD52L2 tumor protein D54|HCCR-binding protein...
562    UHMK1 serine/threonine-protein kinase Kist|KIS...
563    HES5 transcription factor HES-5|class B basic ...
564    ATF7IP activating transcription factor 7-inter...
565    RASSF4 ras association domain-containing prote...
566    BLID BH3-like motif-containing cell death indu...
567    BCL2L14 apoptosis facilitator Bcl-2-like prote...
568    FCRL2 Fc receptor-like protein 2|IFGP family p...
569    ADGRF1 adhesion G-protein coupled receptor F1|...
570    UBE2Q2 ubiquitin-conjugating enzyme E2 Q2|E2 u...
571    FCRL1 Fc receptor-like protein 1|IFGP family p...
572    FHOD3 FH1/FH2 domain-containing protein 3|form...
573    TTC12 tetratricopeptide repeat protein 12|TPR ...
574    TAL2 T-cell acute lymphocytic leukemia protein...
575    CREBL2 cAMP-responsive element-binding protein...
576    CLEC1 C-type lectin-like domain family 1|DC-a...
577    SLX4IP protein SLX4IP
578    HMHB1 minor histocompatibility protein HB-1|mi...
579    ZNF296 zinc finger protein 296|zinc finger pro...
580    MIR520H hsa-mir-520h
581    RCSD1 capZ-interacting protein|RCSD domain-con...
582    LOC107980440 ABL proto-oncogene 1, non-recepto...
583    LOC107963955 BCR p210 Philadelphia chromosome ...
584    LOC107963951 BCR p190 Philadelphia chromosome ...
585    HOXA@ homeo box A cluster
586    E2A single-stranded DNA-binding protein
Length: 573, dtype: object
```

Data Preparation

- Tokenization
- remove special characters and numbers
- remove stopwords
- define methods for bigrams and trigrams

In [4]:

```
# Tokenize
df2_tokens = df2.apply(word_tokenize)
print(df2_tokens)
```

```
0      [IKZF1, DNA-binding, protein, Ikaros|CLL-assoc...
1      [ARID5B, AT-rich, interactive, domain-containi...
2      [CEBPE, CCAAT/enhancer-binding, protein, epsil...
3      [DDC, aromatic-L-amino-acid, decarboxylase|dop...
4      [TSPAN7, tetraspanin-7|CD231, antigen|T-cell, ...
6      [NALT1, MIR4674, host, gene, (, non-protein, c...
8      [CDKN2A, cyclin-dependent, kinase, inhibitor, ...
9      [PDE4B, cAMP-specific, 3',5'-cyclic, phosphodi...
10     [IL15, interleukin-15]
11     [TP63, tumor, protein, 63|amplified, in, squam...
12     [GATA3, trans-acting, T-cell-specific, transcr...
13     [NCOA3, nuclear, receptor, coactivator, 3|CBP-...
14     [RYR2, ryanodine, receptor, 2|cardiac, muscle,...
15     [KCNA3, potassium, voltage-gated, channel, sub...
16     [PIP4K2A, phosphatidylinositol, 5-phosphate, 4...
17     [MAML2, mastermind-like, protein, 2|mam-2|mast...
18     [PTPRJ, receptor-type, tyrosine-protein, phosph...
19     [PARD3, partitioning, defective, 3, homolog|CT...
20     [RXFP1, relaxin, receptor, 1|leucine-rich, rep...
21     [TBXAS1, thromboxane-A, synthase|TXA, synthase...
22     [GRB10, growth, factor, receptor-bound, protei...
23     [ELMO1, engulfment, and, cell, motility, prote...
24     [TP53, cellular, tumor, antigen, p53|antigen, ...
25     [MAGI2, membrane-associated, guanylate, kinase...
26     [GSTT1, glutathione, S-transferase, theta-1|GS...
27     [RIT2, GTP-binding, protein, Rit2|GTP-binding,...
28     [RSP02, R-spondin-2|R-spondin, 2, homolog|roof...
29     [PYGL, glycogen, phosphorylase, ,, liver, form...
30     [MTHFR, methylenetetrahydrofolate, reductase|5...
31     [KCNE4, potassium, voltage-gated, channel, sub...
...
557    [SIAE, sialate, 0-acetylerase|H-Lse|cytosol...
558    [MLXIP, MLX-interacting, protein|MLx, interact...
559    [TSGA10, testis-specific, gene, 10, protein|ca...
560    [DPH1, 2-, (, 3-amino-3-carboxypropyl, ), hist...
561    [TPD52L2, tumor, protein, D54|HCCR-binding, pr...
562    [UHMK1, serine/threonine-protein, kinase, Kist...
563    [HES5, transcription, factor, HES-5|class, B, ...
564    [ATF7IP, activating, transcription, factor, 7-...
565    [RASSF4, ras, association, domain-containing, ...
566    [BLID, BH3-like, motif-containing, cell, death...
567    [BCL2L14, apoptosis, facilitator, Bcl-2-like, ...
568    [FCRL2, Fc, receptor-like, protein, 2|IFGP, fa...
569    [ADGRF1, adhesion, G-protein, coupled, recepto...
570    [UBE2Q2, ubiquitin-conjugating, enzyme, E2, Q2...
571    [FCRL1, Fc, receptor-like, protein, 1|IFGP, fa...
572    [FHOD3, FH1/FH2, domain-containing, protein, 3...
573    [TTC12, tetratricopeptide, repeat, protein, 12...
574    [TAL2, T-cell, acute, lymphocytic, leukemia, p...
575    [CREBL2, cAMP-responsive, element-binding, pro...
576    [CLECL1, C-type, lectin-like, domain, family, ...
577    [SLX4IP, protein, SLX4IP]
578    [HMMB1, minor, histocompatibility, protein, HB...
579    [ZNF296, zinc, finger, protein, 296|zinc, fing...
580    [MIR520H, hsa-mir-520h]
581    [RCSD1, capZ-interacting, protein|RCSD, domain...
582    [LOC107980440, ABL, proto-oncogene, 1, ,, non-...
583    [LOC107963955, BCR, p210, Philadelphia, chromo...
584    [LOC107963951, BCR, p190, Philadelphia, chromo...
585    [HOXA, @, homeo, box, A, cluster]
586    [E2A, single-stranded, DNA-binding, protein]
Length: 573, dtype: object
```

In [5]:

```
# remove special characters and numbers
def prepare_text_for_lda(text):
    cleaned = []
    for word in text:
        for element in word:
            cleaned.append(re.sub('[^a-zA-Z+?0-9_]+', '', element)) # remove all special characters and numbers
            cleaned.append(re.sub(r'^\w', '', element))
        cleaned = list(dict.fromkeys(cleaned)) # remove duplicates
    return cleaned

df2_tokens = prepare_text_for_lda(df2_tokens)
df2_tokens
```


Out[5]:

```
['IKZF1',  
 'DNAbinding',  
 'protein',  
 'IkarosCLLassociated',  
 'antigen',  
 'KW6ikaros',  
 'family',  
 'zinc',  
 'finger',  
 'lymphoid',  
 'transcription',  
 'factor',  
 'LyF1protein',  
 'phosphatase',  
 '1',  
 '',  
 'regulatory',  
 'subunit',  
 '92zinc',  
 'subfamily',  
 '1A',  
 'Ikaros',  
 'ARID5B',  
 'ATrich',  
 'interactive',  
 'domaincontaining',  
 '5BARID',  
 '5BATrich',  
 'domain',  
 '5B',  
 'MRF1like',  
 'proteinmodulator',  
 'recognition',  
 '2',  
 'MRF2',  
 'CEBPE',  
 'CCAATenhancerbinding',  
 'epsilonCCAATenhancer',  
 'binding',  
 'CEBP',  
 'epsilonEBP',  
 'epsilon',  
 'DDC',  
 'aromaticLaminoacid',  
 'decarboxylasedopa',  
 'decarboxylase',  
 'aromatic',  
 'Lamino',  
 'acid',  
 'TSPAN7',  
 'tetraspanin7CD231',  
 'antigenTcell',  
 'acute',  
 'lymphoblastic',  
 'leukemia',  
 'associated',  
 '1cell',  
 'surface',  
 'glycoprotein',  
 'A15membrane',  
 'component',  
 'chromosome',  
 'X',  
 'marker',  
 '1membrane',  
 '1tetraspanin',  
 'proteintransmembrane',  
 '4',  
 'superfamily',  
 '2btransmembrane',  
 'member',  
 '2transmembrane',  
 'A15tspan7',  
 'NALT1',  
 'MIR4674',  
 'host',  
 'gene',  
 'nonprotein',  
 'coding',  
 'MIR4674HG',  
 'LINC01573NOTCH1',  
 'long',  
 'noncoding',  
 'RNA',  
 'in',  
 'pediatric',  
 'TALLlong',  
 'intergenic',  
 '1573',  
 'CDKN2A',
```

'cyclindependent',
'kinase',
'inhibitor',
'2ACDK4',
'p16INK4alternative',
'reading',
'framecell',
'cycle',
'negative',
'regulator',
'betacyclindependent',
'Acyclindependent',
'2A',
'melanoma',
'p16',
'inhibits',
'CDK4',
'multiple',
'tumor',
'suppressor',
'PDE4B',
'cAMPspecific',
'35cyclic',
'phosphodiesterase',
'48duncelike',
'E4phosphodiesterase',
'4B',
'E4',
'dunce',
'homolog',
'Drosophila',
'IL15',
'interleukin15',
'TP63',
'63amplified',
'squamous',
'cell',
'carcinomachronic',
'ulcerative',
'stomatitis',
'proteinkeratinocyte',
'KETtransformationrelated',
'63tumor',
'p53competing',
'GATA3',
'transacting',
'Tcellspecific',
'GATA3GATABinding',
'3',
'NCOA3',
'nuclear',
'receptor',
'coactivator',
'3CBPinteracting',
'proteinamplified',
'breast',
'cancer',
'proteinclass',
'E',
'basic',
'helixloophelix',
'42receptorassociated',
'3steroid',
'3thyroid',
'hormone',
'activator',
'molecule',
'RVR2',
'ryanodine',
'2cardiac',
'muscle',
'receptorcalcium',
'release',
'channelcardiactype',
'receptorislettype',
'receptorkidneytype',
'receptorryanodine',
'cardiac',
'type',
'KCNA3',
'potassium',
'voltagegated',
'channel',
'A',
'3potassium',
'voltage',
'gated',
'shaker',
'related',
'shakerrelated',
'3type',
'n',

'channelvoltagegated',
'K',
'+',
'HuKIIIVoltagegated',
'Kv13voltagegated',
'Kv13',
'PIP4K2A',
'phosphatidylinositol',
'5phosphate',
'4kinase',
'type2',
'alpha1phosphatidylinositol',
'2alpha1phosphatidylinositol4phosphate',
'kinase1phosphatidylinositol4phosphate5kinasePI',
'5',
'P',
'II',
'alphaPIP4KIIalphaPIP5KIIIP5KIIalphaPtdIns',
'P5kinase',
'B',
'isoformdiphosphoinositide',
'2alphaphosphatidylinositol',
'alphaphosphatidylinositol4phosphate',
'5kinase',
'alphaptdIns',
'C',
'isoformptdIns',
'P4kinase',
'isoform',
'2alphatype',
'phosphatidylinositol4phosphate',
'53',
'MAML2',
'mastermindlike',
'2mam2mastermindlike',
'PTPRJ',
'receptortype',
'tyrosineprotein',
'etaCD148',
'antigenDEP1HPTP',
'etaRPTPDensityenhanced',
'1human',
'density',
'enhanced',
'phosphatase1protein',
'tyrosine',
'J',
'polypeptideproteintyrosine',
'etasusceptibility',
'to',
'colon',
'mouse',
'of',
'PARD3',
'partitioning',
'defective',
'homologCTCL',
'se25PAR3alphaatypical',
'PKC',
'isotypespecific',
'interacting',
'proteinbazookapar3',
'polarity',
'alphapar3',
'homologprotein',
'118',
'RXFP1',
'relaxin',
'1leucinerich',
'repeatcontaining',
'G',
'proteincoupled',
'7relaxininsulin',
'like',
'peptide',
'TBXAS1',
'thromboxaneA',
'synthaseTXA',
'synthasecytochrome',
'P450',
'5A1cytochrome',
'polypeptide',
'1platelet',
'cytochrome',
'Vthromboxane',
'synthase',
'platelet',
'GRB10',
'growth',
'receptorbound',
'10GRB10',
'adapter',

'proteinGRB10',
'adaptor',
'proteininsulin',
'receptorbinding',
'GrbIRmaternally',
'expressed',
'ELMO1',
'engulfment',
'and',
'motility',
'Icd12',
'TP53',
'cellular',
'p53antigen',
'NYCO13mutant',
'53p53',
'suppressorphosphoprotein',
'p53transformationrelated',
'53tumor',
'suppressor',
'p53',
'MAGI2',
'membraneassociated',
'guanylate',
'WW',
'PDZ',
'2activin',
'1atrophin1interacting',
'Amembraneassociated',
'inverted',
'GSTT1',
'glutathione',
'Stransferase',
'theta1GST',
'classttheta1glutathione',
'transferase',
'T11',
'RIT2',
'GTPbinding',
'Rit2GTPbinding',
'Roc2Riclike',
'neuronsraslike',
'without',
'CAAX',
'RSP02',
'Rspondin2Rspondin',
'homologroof',
'platespecific',
'spondin2',
'PYGL',
'glycogen',
'phosphorylase',
'liver',
'formphosphorylase',
'MTHFR',
'methylenetetrahydrofolate',
'reductase510methylenetetrahydrofolate',
'reductase',
'NADPH',
'NAD',
'H',
'KCNE4',
'4MINKrelated',
'3cardiac',
'accessory',
'4minimum',
'ion',
'channelrelated',
'beta',
'MiRP3potassium',
'4potassium',
'Iskrelated',
'KCNMB2',
'calciumactivated',
'beta2BK',
'2BK',
'beta2MaxiK',
'subunitMaxiK',
'betasubunit',
'2big',
'2charybdotoxin',
'beta2hCG1646471hbeta2k',
'VCA',
'beta2large',
'conductance',
'subunitlargeconductance',
'Ca2+activated',
'Ca2activated',
'K+',
'beta2',
'subunitpotassium',
'M',

'2potassium',
'large',
'2slobeta2',
'SLCO3A1',
'solute',
'carrier',
'organic',
'anion',
'transporter',
'3A1PGE1',
'transporterorganic',
'polypeptiderelated',
'3organic',
'aniontransporting',
'Dsodiumindependent',
'Dsolute',
'21',
'11solute',
'11',
'DGKB',
'diacylglycerol',
'beta90',
'kDa',
'kinaseDAG',
'betadiacylglycerol',
'90kDadiglyceride',
'ABCB1',
'multidrug',
'resistance',
'1ATPbinding',
'cassette',
'MDRTAP',
'1P',
'glycoproteinPglycoprotein',
'1colchicin',
'sensitivitydoxorubicin',
'KRAS',
'GTPase',
'KRasKRas',
'2Kras',
'p21',
'proteinKirsten',
'rat',
'sarcoma',
'viral',
'oncogene',
'homologKirsten',
'protooncogenePR310',
'cKras',
'oncogeneckirascKirstenras',
'proteincellular',
'cKiras2',
'protooncogenecellular',
'transforming',
'protooncogeneoncogene',
'KRAS2transforming',
'p21vKiras2',
'Kirsten',
'MYRIP',
'rab',
'effector',
'MyRIPSlp',
'homologue',
'lacking',
'C2',
'domainsexophilin8myosinVIIa',
'Rabinteracting',
'proteinslp',
'domains',
'csynaptotagminlike',
'domainscsynaptotagminlike',
'KRT85',
'keratin',
'cuticular',
'Hb5hair',
'K212hard',
'5keratin',
'85',
'IIkeratin',
'hair',
'5type',
'Hb5typeII',
'Kb25',
'MYC',
'myc',
'protooncogene',
'proteinavian',
'myelocytematosis',
'homologclass',
'39mycrelated',
'translationlocalization',
'factorprotooncogene',

'cMyctranscription',
'p64vmc',
'avian',
'homologvmc',
'BCL2',
'apoptosis',
'Bcl2Bcell',
'CLLLymphoma',
'2protein',
'50',
'ERC2',
'ERC',
'2CAZassociated',
'structural',
'proteincytomatrix',
'p110',
'AIG1',
'androgeninduced',
'NPFFR1',
'neuropeptide',
'FF',
'1Gprotein',
'coupled',
'147RFamiderelated',
'OT7T022neuropeptide',
'ST6GALNAC3',
'alphaNacetylgalactosaminide',
'alpha26sialyltransferase',
'3GalNac',
'IIIST6',
'alphaNacetylneuraminy123betagalactosyl13',
'Nacetylgalactosaminide',
'alpha26sialyltranST6',
'3ST6',
'GalNac',
'3alphaNacetylgalactosaminide',
'IIIsialyltransferase',
'7',
'Nacetyl',
'galactosaminide',
'Csialyltransferase',
'7C',
'ZNF230',
'230zinc',
'FDZF2',
'INTS10',
'integrator',
'complex',
'10',
'NOTCH1',
'neurogenic',
'locus',
'notch',
'1Notch',
'translocationassociatednotch',
'1translocationassociated',
'TAN1',
'MAIP1',
'mAAA',
'proteaseinteracting',
'mitochondrial',
'ST8SIA6',
'alpha28sialyltransferase',
'8FST8',
'alphaNacetylneuraminate',
'6sialyltransferase',
'8F',
'alpha2',
'8sialyltransferase',
'sialyltransferase',
'St8Sia',
'VI',
'FLT3',
'FLT3CD135',
'antigenFL',
'cytokine',
'receptorfetal',
'2fmslike',
'3growth',
'IIISTem',
'ABL1',
'ABL1ABL',
'nonreceptor',
'kinaseAbelson',
'1bcrcabl',
'proteincabl',
'kinaseprotooncogene',
'cAblprotooncogene',
'ABL1truncated',
'ABL',
'kinasevabl',
'Abelson',

'murine',
'RUNX1',
'runtrelated',
'1AML1EVI1',
'fusion',
'proteinPEA2alpha',
'BPBP2alpha',
'BSL33',
'enhancer',
'alpha',
'subunitSL3AKV',
'corebinding',
'subunitacute',
'myeloid',
'proteincorebinding',
'runt',
'2oncogene',
'AML1polyomavirus',
'enhancerbinding',
'subunitrunt',
'JAK2',
'JAK2JAK2Janus',
'a',
'TYMS',
'thymidylate',
'synthaseTSase',
'KMT2A',
'histonelysine',
'Nmethyltransferase',
'2ACXXCtype',
'7lysine',
'specific',
'methyltransferase',
'2Alysine',
'2Amixed',
'lineage',
'1myeloidlymphoid',
'or',
'mixedlineage',
'trithorax',
'trithoraxlike',
'proteinzinc',
'HRX',
'TPMT',
'thiopurine',
'SmethyltransferaseSadenosylLmethionine',
'Smethyltransferase',
'GCSAML',
'germinal',
'centerassociated',
'signaling',
'motilitylike',
'MAP1LC3B2',
'microtubuleassociated',
'proteins',
'1A1B',
'light',
'chain',
'2Microtubuleassociated',
'3Blike',
'OR2C3',
'olfactory',
'2C3olfactory',
'2C4olfactory',
'2C5olfactory',
'OR130olfactory',
'4olfactory',
'pseudogene',
'PTEN',
'345trisphosphate',
'3phosphatase',
'dualspecificity',
'PTENMMAC1',
'tensin',
'deleted',
'on',
'10mitochondrial',
'PTENAlphamitochondrial',
'alphamutated',
'advanced',
'cancers',
'1phosphatase',
'tensinlike',
'proteinphosphatidylinositol345trisphosphate',
'PTENprotein',
'LMBRD2',
'LMBR1',
'GSTM1',
'Mu',
'1GST',
'HB',
'4GST',

'classmu',
'1HB',
'4S',
'hydroxyalkyl',
'lyaseglutathione',
'Salkyltransferaseglutathione',
'Saralkyltransferaseglutathione',
'Saryltransferaseglutathione',
'M1',
'ETV6',
'ETV6ETS',
'translocation',
'variant',
'6ETSrelated',
'Tel1TEL1',
'oncogeneets',
'6',
'TEL',
'PAX5',
'paired',
'box',
'Pax5Bcell',
'activatorpaired',
'homeotic',
'5paired',
'5transcription',
'PAX',
'AKT1',
'RACalpha',
'serinethreonineprotein',
'kinaseAKT1mPKB',
'alphaRACPKalphaprotein',
'alphaprotooncogene',
'cAktrac',
'alphaserinethreonine',
'kinasevakt',
'thymoma',
'1vakt',
'oncogenelike',
'CDKN2B',
'BCDK',
'inhibitory',
'proteinCDK4B',
'inhibitorMTS2cyclindependent',
'2B',
'p15',
'kinases',
'proteinmultiple',
'2p14INK4bp14_CDK',
'inhibitorp14_INK4Bp15',
'CDK',
'inhibitorp15INK4bp15_INK4B',
'GSTP1',
'PGST',
'classpiGSTP11deafness',
'Xlinked',
'7epididymis',
'secretory',
'Li',
'22fatty',
'ethyl',
'ester',
'III',
'NRAS',
'NRasNras',
'part',
'4neuroblastoma',
'RAS',
'vras',
'homologneuroblastoma',
'homologtransforming',
'NRasvras',
'neuroblastoma',
'TCF3',
'E2alphaNOL1TCF3',
'fusionVDR',
'repressorclass',
'21helixloophelix',
'HE47immunoglobulin',
'1kappaE2binding',
'factornegative',
'vitamin',
'D',
'response',
'elementbinding',
'proteintranscription',
'E2A',
'immunoglobulin',
'factors',
'E12E47',
'3transcription',
'ITF1vitamin',

'receptorinteracting',
'repressor',
'CRLF2',
'receptorlike',
'2ILXRTSLP',
'receptorcytokine',
'CRL2',
'precusorthymic',
'stromal',
'lymphopoietin',
'receptorthymic',
'stromalderived',
'FBXW7',
'FboxWD',
'7Fbox',
'WD',
'repeat',
'containing',
'E3',
'ubiquitin',
'ligaseFbox',
'WD40',
'archipelago',
'Fbox',
'FBW7Fbox',
'FBX30Fbox',
'SEL10archipelagohomolog',
'elegans',
'sel10',
'SLC19A1',
'folate',
'1intestinal',
'1placental',
'transporterreduced',
'1reduced',
'proteinsolute',
'19',
'NQO1',
'dehydrogenase',
'quinone',
'1DTdiaphoraseNAD',
'1NAD',
'Quinone',
'acceptor',
'oxidoreductase',
'menadione',
'oxireductaseazoreductasediaphorase',
'NADHNADPH',
'b5',
'diaphorase4dioxininducible',
'1menadione',
'reductasephylloquinone',
'reductasequinone',
'NOTCH3',
'3Notch',
'3notch',
'NR3C1',
'glucocorticoid',
'receptorglucocorticoid',
'1nuclear',
'group',
'hGRB',
'54',
'77',
'93',
'BCR',
'breakpoint',
'cluster',
'region',
'proteinBCR',
'RhoGEF',
'activating',
'proteinBCRFGFR1',
'chimera',
'proteinFGFR1BCR',
'proteinbreakpoint',
'regionrenal',
'carcinoma',
'NYREN26',
'CYP1A1',
'1A1aryl',
'hydrocarbon',
'hydroxylasecytochrome',
'P1450',
'dioxininduciblecytochrome',
'form',
'6cytochrome',
'1cytochrome',
'I',
'compoundinducible',
'P450Ccytochrome',
'P450P1flavoproteinlinked',

'monoxygenasexenobiotic',
'monoxygenase',
'TERT',
'telomerase',
'reverse',
'transcriptasetelomerase',
'catalytic',
'subunittelomeraseassociated',
'MTOR',
'mTORFK506',
'12rapamycin',
'2FK506binding',
'complexassociated',
'1FKBPrapamycin',
'proteinFKBP12rapamycin',
'1mammalian',
'target',
'rapamycinmechanistic',
'rapamycin',
'serinethreonine',
'FKBP12',
'1rapamycin',
'FRAP2rapamycin',
'DNMT3A',
'DNA',
'cytosine5',
'3ADNA',
'alphaDNA',
'MTase',
'HsaIIIADNA',
'cytosine',
'3A2',
'TAL1',
'Tcell',
'lymphocytic',
'1Tcell',
'leukemialymphoma',
'5class',
'17stem',
'proteintal1',
'product',
'JAK1',
'PBX1',
'preBcell',
'1homeobox',
'PBX1homeobox',
'PRLpreBcell',
'homeobox',
'WT1',
'Wilms',
'proteinWilms',
'PTPN11',
'11PTP2Cproteintyrosine',
'1Dproteintyrosine',
'2C',
'NFKB1',
'NFkappaB',
'p105',
'subunitDNAbinding',
'KBF1NFkappabetanuclear',
'p50',
'subunitnuclear',
'kappaB',
'kappa',
'Bcells',
'BRAF',
'Braf94',
'Braf',
'proteinBRAF',
'p94',
'BRAF',
'serinethreonineproteinmurine',
'vraf',
'B1protooncogene',
'BRAFvraf',
'Bvraf',
'B1',
'ABCC4',
'resistanceassociated',
'4MRPcMOATrelated',
'ABC',
'transporterbA464I21',
'ATPbinding',
'CFTRMRP',
'canalicular',
'multispecific',
'AFF1',
'AF4FMR2',
'1ALL1fused',
'from',
'proteinmyeloidlymphoid',
'translocated',

```
'2preBcell',
'monocytic',
'partner',
'1protooncogene',
'AF4',
'NUDT15',
'nucleotide',
'triphosphate',
'diphosphatase',
'NUDT15oxodGTPase',
'NUDT15mutT',
'2nucleoside',
'diphosphatelinked',
'another',
'moiety',
'hydrolase',
'15nudid',
'nucleoside',
'diphosphate',
'linked',
'motif',
'15probable',
'78dihydro8oxoguanine',
'triphosphatase',
'NUDT15probable',
'8oxodGTP',
'CXCR4',
'CXC',
'chemokine',
'4CD184',
'antigenLPSassociated',
'3SDF1',
'receptorchemokine',
'4chemokine',
'receptorfusinleukocytederived',
'seven',
'transmembrane',
'receptorlipopolysaccharideassociated',
'3neuropeptide',
'Y',
'Y3neuropeptide',
'Y3',
'receptorseven',
'helix',
'receptorseventransmembranesegment',
'spleenstromal',
'cellderived',
'MDM2',
'ubiquitinprotein',
'ligase',
'Mdm2MDM2',
'ligaseMDM2',
'ligaseMdm2',
'homologMdm2',
'transformed',
'3T3',
'double',
'minute',
'proteindouble',
'human',
'p53binding',
'proteinoncoprotein',
'Mdm2',
'IGH',
'diversity',
'heavy',
'chainsJ',
'joining',
'chainsimmunoglobulin',
'variable',
'regionimmunoglobulin',
'clusterimmunoglobulin',
'groupimmunoglobulin',
'locusimmunoglobulin',
'BCL2L11',
'bcl2like',
'11BCL2like',
'facilitator',
'bcl2',
'mediator',
'deathbcl2',
'Bimbccl2related',
...]
```

In [6]:

```
# define stopwords (genome specific)
stopword_list = stopwords.words('english')
stopword_list += ['sapiens', 'homo', '9606', 'single', 'minus', 'plus', 'homeobox', 'human',
                  'strand', 'nc_001460.1', 'receptor', 'factor', 'subunit', 'kinase', 'class', 'homolog',
                  'member', 'alpha', 'oncogene', 'transcription', 'helix']

print(stopword_list)

['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're", "you've", "you'll", "you'd", 'your', 'yours', 'yourself',
 'imself', 'she', "she's", 'her', 'hers', 'herself', 'it', "it's", 'its', 'itself', 'they', 'them', 'their', 'theirs', 'themselves', 'what',
 'that', "that'll", 'these', 'those', 'am', 'is', 'are', 'was', 'were', 'be', 'been', 'being', 'have', 'has', 'had', 'having', 'do', 'does',
 'and', 'but', 'if', 'or', 'because', 'as', 'until', 'while', 'of', 'at', 'by', 'for', 'with', 'about', 'against', 'between', 'into', 'thru',
 'above', 'below', 'to', 'from', 'up', 'down', 'in', 'out', 'on', 'off', 'over', 'under', 'again', 'further', 'then', 'once', 'here', 'there',
 'all', 'any', 'both', 'each', 'few', 'more', 'most', 'other', 'some', 'such', 'no', 'nor', 'not', 'only', 'own', 'same', 'so', 'than', 'too',
 'l', 'just', 'don', "don't", 'should', "should've", 'now', 'd', 'll', 'm', 'o', 're', 've', 'y', 'ain', 'aren', "aren't", 'couldn', "couldn't",
 'esn't', 'hadn', "hadn't", 'hasn', "hasn't", 'haven', "haven't", 'isn', "isn't", 'ma', 'mightn', "mightn't", 'mustn', "mustn't", 'needn',
 'n', "shouldn't", 'wasn', "wasn't", 'weren', "weren't", 'won', "won't", 'wouldn', "wouldn't", 'sapiens', 'homo', '9606', 'single', 'minus',
 'nc_001460.1', 'receptor', 'factor', 'subunit', 'kinase', 'class', 'homolog', 'member', 'alpha', 'oncogene', 'transcription', 'helix']
```

In [7]:

```
# Build the bigram and trigram models
bigram = gensim.models.Phrases(df2_tokens, min_count=5, threshold=100) # higher threshold fewer phrases.
trigram = gensim.models.Phrases(bigram[df2_tokens], threshold=100)

# Faster way to get a sentence clubbed as a trigram/bigram
bigram_mod = gensim.models.phrases.Phraser(bigram)
trigram_mod = gensim.models.phrases.Phraser(trigram)

# Define functions for stopwords, bigrams, trigrams and Lemmatization
def remove_stopwords(texts):
    return [w for w in texts if not w in stopword_list]

def make_bigrams(texts):
    return [bigram_mod[doc] for doc in texts]

def make_trigrams(texts):
    return [trigram_mod[doc] for doc in texts]
```

Call the above declared methods

In [8]:

```
data_words_nostops = remove_stopwords(df2_tokens)

data_words_bigrams = make_bigrams(data_words_nostops)

data_words_trigrams = make_trigrams(data_words_bigrams)

print(data_words_bigrams)
```

[illegible]

[illegible]

[illegible]

[D, 'h', 'a', 't', '1', '5'], [n, 'u', 'c', 'l', 'e', 'o', 't', 'i', 'd', 'e'], [t, 'r', 'i', 'p', 'h', 'o', 's', 'p', 'h', 'a', 't', 'e'], [n, 'u', 'c', 'l', 'e', 'o', 't', 'i', 'd', 'e'], [N, 'U', 'D', 'T', '1', '5', '8', 'o', 'x', 'o', 'd', 'G', 'T', 'P', 'a', 's', 'e'], [N, 'U', 'D', 'T', '1', '5', '8', 'o', 'x', 'o', 'd', 'G', 'T', 'P', 'a', 's', 'e'], [2, 'n', 'u', 'c', 'l', 'e', 'o', 's', 'i', 'd', 'e'], [d, 'i', 'p', 'h', 'o', 's', 'p', 'h', 'a', 't', 'e', 'l', 'i', 'n', 'k', 'e', 'e', 'r'], [m, 'o', 'i', 'e', 't', 'y'], [h, 'y', 'd', 'r', 'o', 'l', 'a', 's', 'e'], [1, '5', 'n', 'u', 'd', 'i', 'x'], [n, 'u', 'c', 'l', 'e', 'o', 't', 'i', 'd', 'e'], [d, 'i', 'p', 'h', 'o', 's', 'p', 'h', 'a', 't', 'e'], [l, 'i', 'n', 'k', 'e', 'd'], [m, 'o', 't', 'i', 'f'], [1, '5', 'p', 'h', '8', 'd', 'i', 'h', 'y', 'd', 'r', 'o', '8', 'o', 'x', 'o', 'g', 'u', 'a', 'n', 'i', 'n', 'e'], [t, 'r', 'i', 'p', 'h', 'o', 's', 'p', 'h', 'a', 't', 'e'], [N, 'U', 'D', 'T', '1', '5', 'p', 'r', 'o', 'b', 'a', 'b', 'l', 'e'], [8, 'o', 'x', 'o', 'd', 'G', 'T', 'P'], [C, 'X', 'C', 'R', '4', 'm', 'o', 'k', 'i', 'n', 'e'], [4, 'C', 'D', '1', '8', '4'], [a, 'n', 't', 'i', 'g', 'e', 'n', 'L', 'P', 'S', 'a', 's', 's', 'o', 'c', 'S', 'D', 'F', '1'], [r, 'e', 'c', 'e', 'p', 't', 'o', 'r', 'c', 'h', 'e', 'm', 'o', 'k', 'i', 'n', 'e'], [4, 'c', 'h', 'e', 'm', 'o', 'c', 'e', 'p', 't', 'o', 'r', 'f', 'u', 's', 'i', 'n', 'l', 'e', 'u', 'k', 'o', 'c', 'y', 't', 'e', 'd', 'e', 'r', 'i', 'v', 'e', 'd'], [a, 'n', 's', 'm', 'e', 'm', 'b', 'r', 'a', 'n', 'e'], [r, 'e', 'c', 'e', 'p', 't', 'o', 'r', 'l', 'i', 'i', 'p', 'o', 'p', 'o', 'l', 'y', 'i', 'd', 'e', 'a', 's', 's', 'o', 'c', 'i', 'a', 't', 'e', 'd'], [3, 'n', 'e', 'u', 'r', 'o', 'p', 'e', 'p', 't', 'i', 'd', 'e'], [Y, 'o', 'p', 'e', 'p', 't', 'i', 'd', 'e'], [Y, '3'], [r, 'e', 'c', 'e', 'p', 't', 'o', 'r', 's', 'e', 'v', 'e', 'n', 'e', 'n'], [r, 'e', 'c', 'e', 'v', 'e', 'n', 't', 'r', 'a', 'n', 's', 'm', 'e', 'm', 'b', 'r', 'a', 'n', 'e', 's', 'e', 'g', 'm', 'e', 'n', 't'], [s, 'p', 'l', 'e', 'a', 'l'], [c, 'e', 'l', 'l', 'd', 'e', 'r', 'i', 'v', 'e', 'd'], [M, 'D', 'M', '2'], [u, 'b', 'i', 'q', 'u', 'i', 't', 'i', 'n', 'l', 'i', 'g', 'a', 's', 'e'], [M, 'd', 'm', '2', 'M', 'M', '2'], [l, 'i', 'g', 'a', 's', 'e', 'M', 'D', 'M', '2'], [l, 'i', '2], [h, 'i', 'o', 'm', 'o', 'l', 'o', 'g', 'M', 'd', 'm', '2], [t, 'r', 'a', 'n', 's', 'f', 'o', 'r', 'm', 'e', 'd'], [3, 'T', '3], [m, 'i', 'n', 'u', 't', 'e], [p, 'r', 'o', 't', 'e', 'i', 'n', 'd', 'o', 'u', 'b', 'l', 'e], [p, '5', '3', 'b', 'i', 'n', 'd', 'i', 'e', 'i', 'n', 'o', 'n', 'c', 'o', 'p', 'r', 'o', 't', 'e', 'i', 'n'], [M, 'd', 'm', '2], [I, 'G', 'H], [d, 'i', 'v', 'e', 'r', 'v', 'y], [c, 'h', 'a', 'i', 'n', 's', 'J], [j, 'o', 'i', 'n', 'i', 'n', 'g], [c, 'h', 'a', 'i', 'n', 's', 'i', 'm', 'm', 'u', 'i', 'n', 't], [v, 'a', 'r', 'i', 'a', 'b', 'l', 'e], [r, 'e', 'g', 'i', 'o', 'n', 'i', 'm', 'm', 'u', 'n', 'o', 'g', 'l', 'o', 'b', 'u', 'l', 'i', 'n', 't], [g, 'r', 'o', 'u', 'p', 'i', 'm', 'm', 'u', 'l', 'i', 'n', 't], [l, 'o', 'c', 'u', 's', 'i', 'm', 'm', 'u', 'n', 'o', 'g', 'l', 'o', 'b', 'u', 'l', 'i', 'n', 't], [B, 'C', 'L', '2', 'L', 'l', 'i', 'k', 'e], [1, '1', 'B', 'C', 'L', '2', 'l', 'i', 'k', 'e], [f, 'a', 'c', 'i', 'l', 'i', 't', 'a', 't', 'o', 'r], [b, 'a', 't', 'o', 'r], [d, 'e', 'a', 't', 'h], [d, 'e', 'a', 't', 'h], [B, 'i', 'm', 'b', 'c', 'l', '2], [B, 'i', 'm', 'b', 'c', 'l', '2], [e, 'l', 'e', 't', 'e', 'd', 'n', 'n], [d, 'e', 'a', 't', 'h], [a, 'g', 'o', 'n', 'i', 's', 't], [X, 'R', 'C', 'C', '1], [r, 'e', 'p', 'a', 'i', 'r], [X, 'R', 'y], [c, 'o', 'm', 'p', 'l', 'e', 'x], [e, 'm', 'e', 't', 'i', 'n', 'g], [C, 'h', 'i', 'n', 'e', 's', 'e], [h, 'a', 'm', 's', 't', 'e], [1, 'X', 'r', 'a', 'y], [c, 'r', 'o', 's', 's', 'i', 'n', 'g], [c, 'r', 'o', 's', 's', 'i', 'n', 'g], [V, 'E', 'G], [1, 'l', 'a', 'r', 'i], [e, 'n', 'd', 'o', 't', 'h', 'e', 'l', 'i', 'a', 'l], [A, 'v', 'a', 's', 'c', 'u', 'l', 'a', 'r], [A, '1', '2', 'a', 'r], [A, '1', '6', '5', 'v', 'a', 's', 'c', 'u', 'l', 'a', 'r], [p, 'e', 'r', 'm', 'e', 'a', 'b', 'i', 'l', 'i', 't', 'y], [M, 's', 'a', 'm', 'i', 'r', '2', '2', '3], [B, 'T', 'G', '1], [B, 'T', 'G', '1', 'B', 'c', 'e', 'l', 'l], [p, 'r', 'o', 't', 'e', 'i', 'a', 'n', 't', 'i', 'p', 'r', 'o', 'l', 'i', 'f', 'e', 'r', 'a', 't', 'i', 'v', 'e], [L, 'M', 'O', '2], [r, 'h', 'o', 'm', 'b', 'o', 'M], [2, 'T', 'c', 'e', 'l', 'l], [2, 'c', 'y', 's', 't', 'e', 'i', 'n', 'e', 'r', 'i', 'c', 'h], [T, 'T', 'G', '2', 'r', 'h], [l, 'i', 'k', 'e], [N, 'B', 'N], [n, 'i', 'b', 'r', 'i', 'n', 'N], [i, 'j', 'm', 'e', 'g', 'e', 'n'], [b, 'r', 'e', 'a', 'k], [r, 'o', 'm', 'e], [n, 'i', 'b', 'r', 'i', 'n], [p, '9', '5', 'p, '9', '5], [M, 'R', 'E', '1', '1', 'R', 'A', 'D', '5', '0], [n, 's', 'c', 'r', 'i', 'p', 't', 'i', 'o', 'n', 'a', 'l], [E, 'R', 'G', 'E', 'R', 'G], [E, 'T', 'S], [f, 'a', 'c', 't', 'o', 'p, 'r', 'o', 't', 'e', 'i', 'n', 'T', 'M', 'P', 'R', 'S', 'S', '2', 'E', 'R', 'G], [f, 'u', 's', 'i', 'o', 'n', 'e', 'r', 'y', 't, 't', 'r', 'a', 'n', 's', 'f', 'o', 'r', 'm', 'a', 't', 'i', 'o', 'n', 's', 'p', 'e', 'c', 'i', 'f', 'i', 'c], [1, '0', 'e', 't', 'd', 't', 'r', 'a', 'n', 's', 'c', 'r', 'i', 'p', 't', 'i', 'o', 'n', 'a', 'l], [v, 'e', 't', 's], [e, 'r', 'y], [t, 'h', 'r', 'o', 'i', 's], [v, 'i', 'r', 'u', 's], [E, '2', '6], [h, 'o', 'm', 'o', 'l', 'o', 'g', 'v', 'e', 't', 's], [r, 'e', 'l', 'a', 't, 'K', 'I', 'R', '3', 'D', 'S', '1], [K, 'i', 'l', 'l', 'e', 'r], [i, 'm', 'm', 'u', 'n', 'o', 'g', 'l', 'o', 'b', 'u', 'l', 'i', 'n', 't', 'S', '1', 'k', 'i', 'l', 'l', 'e', 'r], [K, 'I', 'R', '3', 'D', 'S', '1', 'k', 'i', 'l', 'l', 'e', 'r], [t, 'h', 'r', 'e', 'e], [t, 'o', 'p', 'l', 'a', 's', 'm', 'i', 'c], [t, 'a', 'i', 'l], [1, 'n', 'a', 't', 'u', 'r', 'a', 'l], [k, 'i', 'l', 'l', 'e', 'a', 't', 'e', 'd], [t, 'r', 'a', 'n', 's', 'c', 'r', 'i', 'p', 't], [T, 'L', 'X', '3], [3, 'h', 'o', 'm', 'e', 'o], [1, 'l', 'o', 'm', 'e', 'o', 'b', 'o', 'x], [H, 'o', 'x', '1', '1', 'L', '2], [H, 'F', 'E], [h, 'e', 'r', 'e', 'd', 'i', 't', 'a', 'r', 'y], [r, 'o', 'm', 'a', 't', 'o', 's', 'i', 's], [p, 'r', 'o', 't', 'e', 'i', 'n', 'M', 'H', 'C], [I, 'l', 'i', 'k', 'e], [H, 'F', 't, 'a', 'r', 'y], [H, 'L', 'A', 'H', 'h', 'i', 'g', 'h], [F, 'e], [R, 'B', '1], [r, 'e', 't', 'i', 'n', 'o', 'b', 'l', 'a', 's', 'o', 'c', 'i', 'a', 't', 'e', 'd], [p, 'r', 'o', 't, 'e', 'i', 'n', 'G', 'O', 'S', '5', '6', '3], [e, 'x', 'o', 'n], [1, 't, 'u', 't', 'i', 'o', 'n], [m, 'u', 't', 'a', 't', 'i', 'o', 'n], [c, 'a', 'u', 's', 'e', 's], [p, 'r', 'e', 'e', 'm', 'a', 't, 'e', 'x', 'o], [G, 'O', 'S', '5', '6', '1], [s, 't', 'o', 'p', 'p, 'r', 'e', 'e', 'p, 'r', 'e', 'e', 't, 'i', 'o', 'b, 'a', 's', 's', 'o', 'c', 'i', 'a', 't',

't', 'i', 'd', 'e', 'n', 'e', 'u', 't', 'r', 'o', 'p', 'h', 'i', 'l', 'a', 'c', 't', 'i', 'v', 'a', 't', 'i', 'n', 'g', '[1', 's', 'm',
'c', 'i', 'b', 'l', 'e', '[8', 't', 'u', 'm', 'o', 'r', '[f', 'a', 'c', 't', 'o', 'r', 'i', 'n', 'd', 'u', 'c', 'e', 'd', '[N', 'C', 'A',
'a', 'l', '[a', 'd', 'h', 'e', 's', 'i', 'o', 'n', '[1', 'a', 'n', 't', 'i', 'g', 'e', 'n', '[r', 'e', 'c', 'o', 'g', 'n', 'i', 'z', 'e',
'l', 'o', 'n', 'a', 'l', '[a', 'n', 't', 'i', 'b', 'o', 'd', 'y', '[5', '1', 'H', '1', 'i', 'n', 'e', 'u', 'r', 'a', 'l', '[N', 'C', 'A',
[h', 'e', 'm', 'a', 't', 'o', 'p', 'o', 'i', 'e', 't', 'i', 'c', '[p', 'r', 'o', 'g', 'e', 'n', 'i', 't', 'o', 'r', '[C', 'D', '3', '4
'R', '1', '[i', 'n', 'a', 'c', 't', 'i', 'v', 'e', '[R', 'O', 'R', '1', 'n', 'e', 'u', 'r', 'o', 't', 'r', 'o', 'p', 'h', 'i', 'c', '[
'r', 'r', 'e', 'l', 'a', 't', 'e', 'd', '[M', 'M', 'P', '2', '[7', '2', '[c', 'o', 'l', 'l', 'a', 't', 'e', 'd', 'n', 'a', 's', 'e', 'c', 'e',
'a', 's', 'e', '[I', 'V', 'A', 'm', 'a', 't', 'r', 'i', 'x', '[7', '2', 'k', 'D', 'a', '[m', 'e', 'g', 'l', 'l', 'o', 'p', 'r', 'i',
'e', '2', 'm', 'a', 't', 'r', 'i', 'x', '[m', 'e', 't', 'a', 'l', 'l', 'o', 'p', 'r', 'o', 't', 'e', 'i', 'n', 'a', 's', 'e', 'I', 'I',
'h', 'i', 'l', '[M', 'A', 'P', 'K', '1', '[m', 'i', 't', 'o', 'g', 'e', 'n', 'a', 'c', 't', 'i', 'v', 'a', 't', 'e', 'd', '[1', 'M', 'A',
[p', '4', '2', 'M', 'A', 'P', 'K', '[2', 'e', 'x', 't', 'a', 'c', 'e', 'l', 'l', 'u', 'l', 'g', 'r', '[s', 'i', 'g', 'n', 'a', 'l',
't', 'e', 'd', '[2', 'm', 'i', 't', 'o', 'g', 'e', 'n', 'a', 'c', 't', 'i', 'v', 'a', 't', 'e', 'd', '[E', 'R', 'K', '2', '[S', 'E',
'l', 'a', 's', 'm', 'i', 'n', 'o', 'g', 'e', 'n', '[1', 'e', 'n', 'd', 'o', 't', 'h', 'e', 'l', 'i', 'a', 'l', '[i', 'n', 'h', 'i', 'b',
'i', 'n', 'e', '[c', 'y', 's', 't', 'e', 'i', 'n', 'e', '[p', 'r', 'o', 't', 'e', 'i', 'n', 'a', 's', 'e', '[c', 'l', 'a', 'd', 'e',
's', 'e', 'r', 'p', 'i', 'n', '[E', '1', 's', 'e', 'r', 'p', 'i', 'n', '[p', 'e', 'p', 't', 'i', 'd', 'a', 's', 'e', '[I', 'L', '7',
'e', 'u', 'k', 'i', 'n', '7', '[a', 'l', 'p', 'h', 'a', 'c', 'D', '1', '2', '7', '[a', 'n', 't', 'i', 'g', 'e', 'n', 'I', 'L', '7', '[
'7', 'R', '[a', 'l', 'p', 'h', 'a', 'i', 'n', 't', 'e', 'r', 'l', 'e', 'u', 'k', 'i', 'n', '[c', 'h', 'a', 'i', 'n', 'i', 'n', 't', 'e',
[n', '[H', '5', '6', '[C', 'D', 'K', 'N', '1', 'A', '[1', 'C', 'D', 'K', 'i', 'n', 't', 'e', 'r', 'l', 'e', 'u', 'k', 'i', 'n', '[1',
'r', 'a', 'c', 't', 'i', 'o', 'n', '[s', 'y', 'n', 't', 'h', 'e', 's', 'i', 's', '[i', 'n', 'h', 'i', 'b', 'i', 't', 'o', 'r', 'c', 'y',
'e', 'n', 'd', 'e', 'n', 't', '[C', 'i', 'p', '1', '[d', 'i', 'f', 'e', 'r', 'e', 'n', 't', 'i', 'a', 't', 'i', 'o', 'n', '[6', 'A,
'e', '[p', '5', '3', 'a', 'c', 't', 'i', 'v', 'a', 't', 'e', 'd', '[f', 'r', 'a', 'g', 'm', 'e', 'n', 't', '[M', 'V', 'B', '[M', 'y',
'n', 'e', '[A', 'M', 'V', 'p', 'r', 'o', 't', 'o', 'o', 'c', 'o', 'g', 'e', 'n', 'e', '[c', 'M', 'y', 'b', 'v', 'm', 'y', 'b', '[r
'a', 's', 't', 'o', 's', 'i', 's', '[S', '1', '0', '0', 'A', '8', '[S', '1', '0', '0', 'A', '8', 'c', 'a', 'l', 'g', 'n', 'a', 'n', 'u
'l', 'p', 'r', 'o', 't', 'e', 'c', 't', 'i', 'n', '[L', '1', 'L', '[s', 'u', 'b', 'u', 'n', 'i', 't', 'c', 'y', 's', 't', 'i', 'c', '[
's', '[a', 'n', 't', 'i', 'g', 'e', 'n', 'l', 'e', 'u', 'k', 'i', 'n', 'c', 'y', 't', 'e', '[L', '1', '[c', 'h', 'a', 'i', 'n', 'm', 'i',
[f', 'a', 'c', 't', 'o', 'r', 'r', 'e', 'l', 'a', 't', 'e', 'd', '[8', 'u', 'r', 'i', 'n', 'a', 'r', 'y', '[s', 't', 'o', 'n', 'e',
'C', 'D', 'A', '[s', 'i', 'n', 'g', 'l', 'e', 's', 't', 'r', 'a', 'a', 'n', 'd', 'e', 'd', '[d', 'e', 'a', 'm', 'i', 'n', 'a', 's', 'e', 'c
'e', '[a', 'm', 'i', 'n', 'o', 'h', 'y', 'd', 'r', 'o', 'l', 'a', 's', 'e', 'e', 'p', 'i', 'd', 'i', 'd', 'y', 'm', 'i', 's', '[2', '8
'r', 'a', 't', 'e', 'd', '[B', 'u', 'r', 'k', 'i', 't', 't', '[1', 'i', 'n', 'e', '[R', 'a', 'm', 'o', 's', '[C', 'X', 'C', 'L', '1
'o', 'k', 'i', 'n', 'e', '[1', '2', 'i', 'n', 't', 'e', 'r', 'c', 'r', 'i', 'n', 'e', '[r', 'e', 'd', 'u', 'c', 'e', 'd', '[h', 'e',
'p', 'r', 'e', 'B', '[g', 'r', 'o', 'w', 't', 'h', 's', 't', 'i', 'm', 'u', 'l', 'a', 't', 'i', 'n', 'g', '[S', 'T', 'A', 'T', '5', 'B
[t', 'n', 'a', 'n', 's', 'd', 'u', 'c', 'e', 'r', '[5', 'B', 't', 'r', 'a', 'n', 's', 'c', 'r', 'i', 'p', 't', 'i', 'o', 'n', '[R',
'F', 'k', 'a', 'p', 'p', 'a', 'B', '[p', '6', '5', 'd', 'e', 'l', 't', 'a', '3', 'N', 'F', 'k', 'a', 'p', 'p', 'a', 'B', '[p', '6', '5
'n', '[p', '6', '5', '[3', 'v', 'n', 'e', 'l', '[r', 'e', 't', 'i', 'c', 'u', 'l', 'l', 'o', 'e', 'n', 'd', 'o', 't', 'h', 'e', 'l', 'i',
'e', 'l', '[H', 'L', 'A', 'G', '[H', 'L', 'A', '[h', 'i', 's', 't', 'o', 'c', 'o', 'm', 'p', 'a', 't', 'i', 'i', 'l', 'i', 't', 'y,
'n', 't', 'i', 'g', 'e', 'n', 'H', 'L', 'A', 'G', '[G', 'M', 'H', 'C', '[I', 'b', '[a', 'n', 't', 'i', 'g', 'e', 'n', 'b', '2', '[m
'b', 'u', 'l', 'i', 'i', 'n', 'm', 'u', 't', 'a', 'n', 't', '[M', 'H', 'C', '[a', 'n', 't', 'i', 'g', 'e', 'n', 'm', 'u', 't', 'a', 'n', 't
[W', 'i', 'l', 'l', 'e', 'b', 'r', 'a', 'n', 'd', '[f', 'a', 'c', 't', 'o', 'r', 'c', 'o', 'a', 'g', 'u', 'l', 'a', 't', 'i', 'o', 'n,
'X', 'A', 'S', '2', '[H', 'O', 'X', 'A', '[a', 'n', 't', 'i', 's', 'e', 'n', 's', 'e', '[P', 'A', 'R', 'P', '1', '[p', 'o', 'l
'b', 'o', 's', 'e', '[p', 'o', 'l', 'y', 'm', 'e', 'r', 'a', 's', 'e', '[1', 'A', 'D', 'P', 'r', 'i', 'b', 'o', 's', 'y', 'l', 't', 'r
'a', 's', 'e', '[N', 'A', 'D', '+', '[A', 'D', 'P', 'r', 'i', 'b', 'o', 's', 'y', 'l', 't', 'r', 'a', 'n', 's', 'f', 'e', 'r', 'a', 's
'h', 'e', 'r', 'i', 'a', '[t', 'o', 'x', 'i', 'l', 'i', 'c', 'e', '[P', 'A', 'R', 'P', '1', 'N', 'A', 'D', '[1', 'p', 'o', 'l,
't', 'a', 's', 'e', 'p', 'o', 'l', 'y', '[A', 'D', 'P', 'r', 'i', 'b', 'o', 's', 'y', 'l', '[t', 'r', 'a', 'n', 's', 'f', 'e', 'r', 'a',
[1', 'p', 'r', 'o', 't', 'e', 'i', 'n', '[p', 'o', 'l', 'y', 'A', 'D', 'P', 'r', 'i', 'b', 'o', 's', 'y', 'l', 't', 'r', 'a', 'n', 's',
[K', 'I', 'R', '2', 'D', 'S', '3', '[2', 'D', 'S', '3', 'M', 'H', 'C', '[N', 'K', '[r', 'e', 'c', 'e', 'p', 't', 'o', 'r', 'N', 'K,
'e', 'r', '[t', 'w', 'o', '[3', 'n', 'a', 't', 'u', 'r', 'a', 'l', '[r', 'e', 'c', 'e', 'p', 't', 'o', 'r', 'n', 'a', 't', 'u', 'r,
[J', 'A', 'K', '3', 'J', 'a', 'n', 'u', 's', '[1', 'e', 'u', 'k', 'i', 'n', 'c', 'y', 't', 'e', '[J', 'a', 'n', 'u', 's', '[C', 'D', 'K,
'y', 'c', 'l', 'i', 'n', 'd', 'e', 'p', 'e', 'n', 'd', 'e', 'n', 't', '[1', 'B', '[p', '2', '7', '[K', 'i', 'p', '1', '[A', 'B', 'C
[t', 'r', 'a', 'n', 's', 'p', 'o', 'r', 't', 'e', 'r', 'A', 'T', 'P', 'i', 'n', 'd', 'i', 'n', 'g', '[G', '2', 'A', 'T', 'P', 'b
[W', 'H', 'I', 'T', 'E', '[J', 'u', 'n', 'i', 'o', 'r', '[b', 'l', 'o', 'o', 'd', '[p', 'r', 'o', 't', 'e', 'i', 'n', 'm', 'i', 't,
'n', 'e', '[p', 'r', 'o', 't', 'e', 'i', 'n', 'm', 'u', 'l', 't', 'i', '[d', 'r', 'u', 'g', '[e', 'f', 'f', 'l', 'u', 'x', '[t', 'r
[2', 'p', 'l', 'a', 'c', 'e', 'n', 't', '[M', 'D', 'R', '[p', 'r', 'o', 't', 'e', 'i', 'n', 'p', 'l', 'a', 'c', 'e,
'f', 'i', 'c', '[t', 'r', 'a', 'n', 's', 'p', 'o', 'r', 't', 'e', 'r', 'u', 'r', 'a', 't', 'e', '[e', 'x', 'p', 'o', 'r', 't', 'e', 'r
D', '[6', 'P', 'H', 'D', 'l', 'i', 'k', 'e', '[p', 'r', 'o', 't', 'e', 'i', 'n', 'c', 'e', 'n', 't', 'r', 'o', 'm', 'e', 'r', 'e', '[
[i', 'n', 't', 'e', 'r', 'l', 'e', 'u', 'k', 'i', 'n', '1', '[a', 'n', 't', 'a', 'g', 'o', 'n', 'i', 's', 't', '[p', 'r', 'o', 't,
'n', 'h', 'i', 'b', 'i', 't', 'o', 'r', 'i', 'n', 't', 'a', 'c', 'e', 'l', 'l', 'u', 'l', 'a', 'r', '[I', 'L', '1', '[I', 'I', 'i
'l', 'l', 'u', 'l', 'a', 'r', '[i', 'c', 'I', 'L', '1', 'n', 'a', '[U', 'G', 'T', '1', 'A', '1', '[U', 'D', 'P', 'g', 'l', 'u', 'c,
'l', 't', 'r', 'a', 'n', 's', 'f', 'e', 'r', 'a', 's', 'e', '[1', '1', 'U', 'D', 'P', '[g', 'l', 'u', 'c', 'u', 'r', 'o', 'n', 's', 's
's', 't', 'e', 'n', 'a', 's', 'e', '[A', '1', 'U', 'D', 'P', '[g', 'l', 'u', 'c', 'o', 's', 'y', 'l', 't', 'r', 'a', 'n', 's', 'f,
'U', 'D', 'P', 'g', 'l', 'u', 'c', 'u', 'r', 'o', 'n', 's', 's', 'y', 'l', 't', 'r', 'a', 'n', 's', 'f', 'e', 'r', 'a', 's', 'e', '[1,
'c', 'u', 'r', 'o', 'n', 's', 'y', 'l', 't', 'r', 'a', 'n', 's', 'f', 'e', 'r', 'a', 's', 'e', '[1', 'A', '1', 'U', 'G', 'T', '1,
'G', 'T', '1', '0', '1', 'U', 'G', 'T', '1', 'b', 'i', 'l', 'i', 'r', 'u', 'b', 'i', 'n', '[U', 'D', 'P', 'g', 'l', 'u', 'c', 'u,
't', 'r', 'a', 'n', 's', 'e', 'r', 'a', 's', 'e', 'b', 'i', 'l', 'i', 'r', 'u', 'b', 'i', 'n', '[1', '1', 'b', 'i', 'l', 'i', 'r', 'u,
'z', 'y', 'm', 'e', '[1', 'b', 'i', 'l', 'i', 'r', 'u', 'b', 'i', 'n', 's', 'p', 'e', 'c', 'i', 'f', 'i', 'c', '[U', 'D', 'P', 'G', 'T
'i', 's', 't', 'o', 'n', 'e', '[a', 'c', 'e', 't', 'y', 'l', 't', 'r', 'a', 'n', 's', 'f', 'e', 'r', 'a', 's', 'e', '[p', '3', '0', '0
'c', 'i', 'a', 't', 'e', 'd', '[p', '3', '0', '0', '1', 'A', 'b', 'i', 'n', 'd', 'i', 'n', 'g', '[3', '0', '0', 'k', 'D', 'h', 'i
'u', 't', 'y', 'r', 'y', 'l', 't', 'r', 'a', 'n', 's', 'f', 'e', 'r', 'a', 's', 'e', '[p', '3', '0', '0', 'h', 'i', 'i', 's', 't', 'o', 'n,
'n', 'y', 'l', 't', 'r', 'a', 'n', 's', 'f', 'e', 'r', 'a', 's', 'e', '[p', '3', '0', '0', 'p', '3', '0', '0', '[H', 'A', 'T', 'p', 'r
'r', 'o', 'p', 'i', 'o', 'n', 'y', 'l', 't', 'r', 'a', 'n', 's', 'f', 'e', 'r', 'a', 's', 'e', '[p', '3', '0', '0', '[E', 'R', 'C', 'C
'a', 'l', '[I', 'I', 'H', '[h', 'e', 'l', 'i', 'c', 'a', 's', 'e', '[X', 'P', 'D', 'B', 'T', 'F', '2', '[p', '8', '0', 'C', 'X', 'P
'c', 'i', 's', 'i', 'n', '[E', 'R', 'C', 'C', '2', 'D', 'N', 'A', '[X', 'P', 'D', '[c', 'e', 'l', 'l', 's', 'T', 'F', 'I', 'I,
'u', 'n', 'i', 't', 'T', 'F', 'I', 'I', 'H', '[b', 'a', 's', 'a', 'l', '[X', 'P', 'B', '[p', '8', '0', 'T', 'F', 'I', 'I', 'H', '[X
[C', '[s', 'u', 'b', 'u', 'n', 'i', 't', 'e', 'x', 'c', 'i', 's', 'i', 'o', 'n', '[c', 'r', 'o', 's', 's', 'c', 'o', 'm', 'p', 'l', 'e
'i', 'o', 'n', '[2', 'e', 'x', 'c', 'i', 's', 'i', 'o', 'n', '[r', 'o', 'd', 'e', 'n', 't', '[d', 'e', 'f', 'i', 'c', 'i', 'e', 'n',
'e', 'm', 'e', 'n', 't', 'a', 't', 'i', 'o', 'n', '[2', 'x', 'e', 'r', 'o', 'd', 'e', 'r', 'm', 'a', '[p', 'i', 'g', 'm', 'e', 'n', 't
'm', 'p', 'l', 'e', 'm', 'e', 'n', 't', 'a', 't', 'i', 'o', 'n', '[D', 'x', 'e', 'r', 'o', 'd', 'e', 'r', 'm', 'a', '[D', 'c', 'o', 'm', 'p', 'l
'n', 'g', '[N', 'T', '5', 'C', '2', '[c', 'y', 't', 'o', 's', 'o', 'l', 'i', 'c', '[p', 'u', 'r', 'i', 'n', 'e', '[5', 'n', 'u', 'c
's', 'e', '5', 'n', 'u', 'c', 'l', 'e', 'o', 't', 'i', 'd', 'a', 's', 'e', '[B', 'I', 'M', 'P', 's', 'p', 'e', 'c', 'i', 'f', 'i', 'c,
'i', 'd', 'y', 'm', 'i', 's', '[s', 'p', 'e', 'r', 'm', '[N', 'A', 'T', '2', '[a', 'n', 'y', 'l', 'a', 'm', 'i', 'n', 'e', '[N', 'a,
'a', 'n', 's', 'f', 'e', 'r', 'a', 's', 'e', '[2', 'N', 'a', 'c', 'e', 't', 'y', 'l', 't', 'r', 'a', 'n', 's', 'f', 'e', 'r', 'a', 's,
'a', 'm', 'i', 'd', 'e', '[a', 'c', 'e', 't', 'y', 'l', 'a', 's', 'e', '[I', 'T', 'P', 'A', '[i', 'n', 'o', 's', 'i', 'n', 'e', '[p
'p', 'h', 'a', 's', 'e', 'e', 'p', 'i', 'd', 'i', 'd', 'y', 'm', 'i', 's', '[p', 'r', 'o', 't', 'e', 'i', 'n', 'i', 'o',
'o', 'p', 'h', 'o', 's', 'p', 'h', 'a', 't', 'a', 's', 'e', '[p', 'y', 'h', 'o', 'p', 'h', 'o', 's', 'p', 'h', 'o', 'y', 'd', 'r,
'n', 'c', 'a', 'n', 'o', 'n', 'i', 'c', 'a', 'l', '[N', 'T', 'P', '[p', 'y', 'r', 'o', 'p', 'h', 'o', 's', 'p', 'h', 'a', 't', 'a', 's
'a', 'n', 'd', 'a', 't', '[p', 'y', 'r', 'o', 'p', 'h', 'o', 's', 'p', 'h', 'a', 't', 'a', 's', 'e', 'n', 'u', 'c', 'l', 'e', 'o,
'p', 'h', 'o', 's', 'p', 'h', 'a', 't', 'e', '[d', 'i', 'p', 'h', 'o', 's', 'p', 'h', 'a', 't', 'a', 's', 'e', 'p', 'u', 't', 'a', 't,
'1', '4', '0', '6', 'P', '[C', 'Y', 'P', '3', 'A', '5', '[3', 'A', '5', 'a', 'r', 'y', 'l', '[H', 'L', 'p', '2', 'c', 'y', 't', 'o,
'c', 'y', 't', 'o', 'c', 'h', 'n', 'o', 'm', 'e', '[I', 'I', 'I', 'A', '[n', 'i', 'p', 'h', 'e', 'd', 'i', 'p', 'i', 'n', 'e', '[o', 'n
'4', '5', '0', 'P', 'C', 'N', '3', 'f', 'l', 'a', 'v', 'o', 'p', 'r', 'o', 't', 'e', 'i', 'n', 'l', 'i', 'n', 'k', 'e', 'd', '[m', 'o,
'n', 'a', 's', 'e', 'm', 'i', 'c', 'r', 'o', 's', 'o', 'm', 'a', 'l', '[L', 'E', 'P', 'R', '[1', 'e', 'p', 't', 'i', 'n', '[r', 'e,
'B', '[r', 'e', 'c', 'e', 'p', 't', 'o', 'r', 'h', 'u', 'B', '2', '1', '9', '[O', 'G', 'G', '1', '[N', 'g', 'l', 'y', 'c', 'o', 's,
'A', '[1', 'y', 'a', 's', 'e', '8', 'h', 'y', 'd', 'r', 'o', 'x', 'y', 'g', 'u', 'a', 'n', 'i', 'n', 'e', '[g', 'l', 'y', 'c', 'o', 's

'P'], ['l', 'y', 'a', 's', 'e', 'D', 'N', 'A', 'a', 'p', 'u', 'r', 'i', 'n', 'i', 'c'], ['a', 'p', 'y', 'r', 'i', 'm', 'i', 'd', 'i', 'n
[l', 'y', 'a', 's', 'e', 'O', 'G', 'G', '1'], [l', 'f'], [C', 'X', 'C', 'L', '1', '0'], [l', '0', '1', '0'], [i', 'n', 't', 'e', 'r
a', 'm', 'm', 'a', 'i', 'n', 'd', 'u', 'c', 'e', 'd'], [l', '0', 'g', 'a', 'm', 'm', 'a', 'a'], [I', 'P', '1', '0', 'i', 'n', 't', 'e', 'r
n', 'd', 'u', 'c', 'i', 'b', 'l', 'e'], [I', 'P', '1', '0', 'p', 'r', 'o', 't', 'e', 'i', 'n'], [g', 'a', 'm', 'm', 'a', 'a'], [i', 'n', 't', 'e', 'r
n', 'e', 's', 'm', 'a', 'l', 'l', 'l'], [C', 'y', 's', 'X', 'C', 'y', 's'], [l', '0', 's', 'm', 'a', 'l', 'l', 'l', 'i', 'n', 'd', 'u', 'c', 'i
[P', 'R', 'O', 'M', '1'], [p', 'r', 'o', 'm', 'i', 'n', 'i', 'n', 'l', 'a', 'n', 't', 'i', 'g', 'e', 'n'], [A', 'C', '1', '3', '3', 'i
i', 'n', 'h', 'e', 'm', 'a', 't', 'o', 'p', 'o', 'i', 'e', 't', 'i', 'c'], [s', 't', 'e', 'm'], [a', 'n', 't', 'i', 'g', 'e', 'n', 'n', 'p
n', 'l', 'i', 'i', 'k', 'e'], [A', 'P', 'E', 'X', '1'], [a', 'p', 'u', 'r', 'i', 'n', 'i', 'c'], [l', 'y', 'a', 's', 'e', 'A', 'P'], [e
e', 'a', 's', 'e'], [I', 'A', 'P'], [l', 'y', 'a', 's', 'e', 'A', 'P', 'E', 'X'], [n', 'u', 'c', 'l', 'e', 'a', 's', 'e'], [m', 'u',
t', 'i', 'o', 'n', 'a', 'l'], [e', 'n', 'z', 'y', 'm', 'e'], [l', 'a', 'p', 'u', 'r', 'i', 'n', 'i', 'c', 'a', 'p', 'y', 'r', 'i', 'b', 'm
[a', 'b', 'a', 's', 'i', 'l', 'e], [e', 'n', 'd', 'o', 'n', 'u', 'c', 'l', 'e', 'a', 's', 'e', 'd', 'e', 'o', 'x', 'y', 'r', 'i', 'b', 'o
e], [R', 'E', 'F', '1', 'r', 'e', 'd', 'o', 'x'], [f', 'a', 'c', 't', 'o', 'r', '1'], [R', 'A', 'D', '5', '1'], [l', 'B', 'R', 'C',
c', 'o', 'n', 't', 'a', 'i', 'n', 'i', 'n', 'g'], [5', 'R', 'A', 'D', '5', '1'], [A', 'R', 'e', 'c', 'A'], [c', 'o', 'l', 'i', 'i], [o
k', 'e], [p', 'r', 'o', 't', 'e', 'i', 'n', 'r', 'e', 'c', 'o', 'm', 'b', 'i', 'n', 'a', 't', 'i', 'o', 'n'], [C', 'R', 'E', 'B', 'i', 'n
[A', 'M', 'P', 'r', 'e', 's', 'p', 'o', 'n', 's', 'i', 'v', 'e], [l', 'a', 'c', 't', 'i', 'v', 'e], [C', 'R', 'E', 'B', 'c', 'A', 'i
n', 's', 'e], [p', 'r', 'o', 't', 'e', 'i', 'n', '1', 'c', 'y', 'c', 'l', 'i', 'c], [a', 'd', 'e', 'n', 'o', 's', 'i', 'n', 'e], [i
h', 'o', 's', 'p', 'h', 'a', 't', 'e], [e', 'l', 'e', 'm', 'e', 'n', 't], [p', 'r', 'o', 't', 'e', 'i', 'n', 'c', 'y', 'c', 'l', 'i
n', 'a', 'n', 's', 'a', 'c', 't', 'i', 'v', 'a', 't', 'o', 'n'], [M', 'I', 'R', '1', '0', '1', '1], [h', 's', 'a', 'm', 'i', 'r', '1
T', '1], [s', 'e', 'r', 'i', 'n', 'e], [h', 'y', 'd', 'r', 'o', 'x', 'y', 'm', 'e', 't', 'h', 'y', 'l', 't', 'r', 'a', 'n', 's', 'f
y', 't', 'o', 's', 'o', 'l', 'i', 'c', 'c', 'y', 't', 'o', 'p', 'l', 'a', 's', 'm', 'i', 'c], [h', 'y', 'd', 'r', 'o', 'x', 'y', 'm',
a', 'n', 's', 'f', 'e', 'r', 'a', 's', 'e', 'g', 'l', 'y', 'c', 'i', 'n', 'e], [h', 'y', 'd', 'r', 'o', 'x', 'y', 'm', 'e', 't', 'h,
f', 'e', 'r', 'a', 's', 'e', 's', 'e', 'r', 'i', 'n', 'e], [s', 'o', 'l', 'u', 'b', 'l', 'e], [m', 'e', 't', 'h', 'y', 'l', 'a', 's
n', 'i', 'c], [h', 'e', 'd', 'g', 'e', 'h', 'o', 'g'], [p', 'r', 'o', 't', 'e', 'i', 'n', 's', 'h', 'h], [u', 'n', 'p', 'r', 'o', 't
t', 'e', 'r', 'm', 'i', 'n', 'a', 'l], [C', 't', 'e', 'r', 'm', 'i', 'n', 'a', 'l], [a', 'u', 't', 'o', 'p', 'r', 'o', 't', 'e', 'i', 'n', 's
m', 'a', 'i', 'n', 's', 'o', 'n', 'i', 'c], [K', 'I', 'R', '2', 'D', 'S', '2'], [2', 'D', 'S', '2', 'C', 'D', '1', '5', '8], [l
l', 'i', 'k', 'e], [J', 'K', 'I', 'R', '2', 'D', 'S', '2'], [K', 'i', 'l', 'l', 'e', 'r', 'c', 'e', 'l', 'l], [I', 'm', 'm', 'u,
l', 'i', 'n', 'l', 'i', 'k', 'e], [R', 'e', 'c', 'e', 'p', 't', 'o', 'r', 'M', 'H', 'C], [r', 'e', 'c', 'e', 'p', 't', 'o', 'r', 'N
t', 'I', 'K', 'i', 'l', 'l', 'e', 'r], [K', 'I', 'R', '2', 'D', 'S', '2', 'i', 'l', 'l', 'e', 'r], [2', 'i', 'l', 'l', 'e
g', 'l', 'i', 'k', 'e], [r', 'e', 'c', 'e', 'p', 't', 'o', 'r', 'k', 'i', 'l', 'l', 'e', 'r', 'c', 'e', 'l', 'l], [p', 'r', 'o', 't
u', 'r', 'a', 'l], [5', 'n', 'a', 't', 'u', 'r', 'a', 'l], [r', 'e', 'c', 'e', 'p', 't', 'o', 'r', 'p', '5', '8], [K', 'I', 'R,
a', 't', 'u', 'r', 'a', 'l], [c', 'l', 'o', 'n', 'e], [C', 'L', '4', '9'], [C', 'B', 'L'], [C', 'B', 'L', 'C', 'a', 's', 'B', 'r,
o', 't', 'i', 'c], [I', 'r', 'e', 't', 'r', 'o', 'v', 'i', 'r', 'a', 'l], [s', 'e', 'q', 'u', 'e', 'n', 'c', 'e', 'C', 'b', 'l', 'l], [l,
N', 'G'], [5', 'S', 'R', 'I', 'N', 'G', 't', 'y', 'p', 'e], [C', 'B', 'L', 'c', 'a', 's', 'i', 't', 'a', 's], [B', 'l', 'i', 'n,
t', 'o', 'o', 'n', 'c', 'o', 'g', 'e', 'n', 'e', 'f', 'r', 'a', 'g', 'i', 'l', 'e], [f', 'o', 'l', 'i', 'c], [r', 'a', 'r', 'e], [C
[C', 'B', 'L', '2', 'p', 'r', 'o', 't', 'o', 'n', 'c', 'o', 'g', 'e', 'n', 'e], [c', 'b', 'l', 's', 'i', 'g', 'n', 'a', 'l]
u', 'c', 't', 'i', 'o', 'n], [T', 'N', 'F', 'S', 'F', '1', '3', 'B], [1', '3', 'B', 'A', 'p', 'o', 'l', 'l], [T', 'A', 'L', 'L', '1,
t', 'i', 'v', 'a', 't', 'i', 'n', 'g], [f', 'a', 'c', 't', 'o', 'r', 'B', 'l', 'y', 'm', 'p', 'h', 'o', 'c', 'y', 't', 'e], [s', 't
o', 'r', 'D', 'e', 'l', 't', 'a', '4], [B', 'A', 'F', 'T', 'N', 'F], [A', 'p', 'o', 'l', 'l', 'a', 't', 'e', 'd], [i
i', 'v', 'a', 't', 'e', 's], [a', 'p', 'o', 't', 'o', 's', 'i', 's', 'd', 'e', 'l', 't', 'a], [B', 'A', 'F', 'd', 'e', 'n
[T', 'N', 'F', '1', 'i', 'k', 'e], [m', 'o', 'l', 'e', 'c', 'u', 'l', 'e', 'e', 'p', 'i', 'd', 'i', 'd', 'y', 'm', 'i', 's], [p', 'r
u', 'm', 'o', 'r', 'l], [l', '3', 'b', 't', 'u', 'm', 'o', 'r], [2', '0', 't', 'u', 'm', 'o', 'r], [7', 'A', 't', 'u', 'm', 'o', 'r', 'r],
i', 'k', 'e], [Z', 'T', 'N', 'F', '4], [I', 'L', '1', '0], [i', 'n', 't', 'e', 'r', 'l', 'e', 'n', 'i', 'n', 't', 'e', 'r], [X', 'R', 'C
t', 'o', 'r', 'c', 'y', 't', 'o', 'k', 'i', 'n', 'e], [H', 'I', 'F', '1', 'A], [h', 'y', 'p', 'o', 'x', 'i', 'a', 'i', 'n', 'd', 'u
a', 'l', 'p', 'h', 'a', 'A', 'R', 'N', 'T], [p', 'r', 'o', 't', 'e', 'i', 'n', 'P', 'A', 'S], [8', 'b', 'a', 's', 'i', 'c', 'h', 'e
p', 'h', 'e', 'l', 'i', 'x', 'P', 'A', 'S], [M', 'O', 'P', '1', 'c', 'l', 'a', 's', 's], [7', '8', 'h', 'y', 'p', 'o', 'x', 'i', 'a
t', 'h', 'y', 'p', 'o', 'x', 'i', 'a], [I', '3', 'h', 'y', 'p', 'o', 'x', 'i', 'a', 'i', 'n', 'd', 'u', 'c', 'i', 'b', 'l', 'e], [f
a', 'l', 'p', 'h', 'a', 'm', 'e', 'm', 'b', 'e', 'r], [P', 'A', 'S], [1', 'm', 'e', 'm', 'b', 'e', 'r], [C', 'E', 'B', 'P', 'A],
A', 'A', 'T', 'e', 'n', 'h', 'a', 'n', 'c', 'e', 'r], [I', 'D', '4], [I', 'D', '4', 'c', 'l', 'a', 's', 's], [2', '7', 'i', 'n', 't
[d', 'o', 'm', 'i', 'n', 'a', 'n', 't], [p', 'r', 'o', 't', 'e', 'i', 'n', 'i', 'n', 'h', 'i', 'b', 'i', 't', 'o', 'r], [X', 'R', 'C
3', 'X', 'r', 'a', 'y], [3', 'X', 'r', 'a', 'y], [C', 'H', 'E', 'K', '1], [C', 'h', 'k', '1', 'C', 'H', 'K', '1], [c', 'h', 'e
t], [h', 'o', 'm', 'o', 'l', 'o', 'g', 'C', 'h', 'e', 'c', 'k', 'p', 'o', 'i', 'n', 't], [S], [p', 'o', 'm', 'b', 'e], [l', 'C
l], [H', 'L', 'A', 'D', 'R', 'B', '1], [m', 'a', 'j', 'o', 'r], [D', 'R], [l', 'H', 'L', 'A], [D', 'R', '1], [c', 'h', 'a,
A', 'D', 'R], [c', 'h', 'a', 'i', 'n', 'h', 'u', 'm', 'a', 'n], [l', 'e', 'u', 'c', 'o', 'c', 'y', 't', 'e], [D', 'R', 'B', '1',
y', 't', 'e], [D', 'R', 'B', '1], [A', 'T', 'R], [A', 'T', 'R', 'F', 'R', 'A', 'P', 'r', 'e', 'l', 'a', 't', 'e', 'd], [p', 'r
E', 'C', '1], [m', 'i', 't', 'o', 's', 'i', 's], [e', 'n', 't', 'r', 'y], [h', 'o', 'm', 'o', 'l', 'o', 'g', 'a', 't', 'a', 'x',
g', 'i', 'e', 'c', 't', 'a', 's], [R', 'a', 'd', '3', 'r', 'e', 'l', 'a', 't', 'e', 'd], [B', 'D', 'N', 'F], [b', 'r,
v', 'e', 'd], [n', 'e', 'u', 'r', 'o', 't', 'r', 'o', 'p', 'h', 'i', 'c], [f', 'a', 'c', 't', 'o', 'r', 'a', 'b', 'n', 'i', 'n', 'e
u', 'r', 'o', 't', 'r', 'o', 'p', 'h', 'i', 'n], [T', 'N', 'F', 'A', 'I', 'P', '3], [a', 'l', 'p', 'h', 'a', 'i', 'n', 'd', 'u', 'c
[7', 'C', 'p', 'u', 't', 'a', 't', 'i', 'v', 'e], [A', '2', '0', 't', 'u', 'm', 'o', 'r], [3', 'z', 'i', 'n', 'c], [A', '2', '0'
l], [1', 'B', '1', 'O', 'A', 'T', 'P', '2', 'l', 'i', 'v', 'e', 'r', 's', 'p', 'e', 'c', 'i', 'f', 'i', 'c], [1', 's', 'o', 'd', 'i
p', 'e', 'n', 'd', 'e', 'n', 't], [2', 's', 'o', 'l', 'u', 't', 'e], [B', 'C', 'L', '6], [B', 'c', 'e', 'l', 'l], [p', 'r', 'o
B', 'c', 'e', 'l', 'l], [t', 'r', 'a', 'n', 's', 'c', 'r', 'i', 'p', 't', 'B', 'C', 'L', '5', 'B', 'C', 'L', '6', 'c', 'y', 's', 'h,
t', 'o', 'r', 'l', 'y', 'm', 'p', 'h', 'o', 'm', 'a', 'a', 's', 's', 'o', 'c', 'i', 'a', 't', 'e', 'd], [3', 'p', 'r', 'o', 't', 'e,
z', 'i', 'n', 'c], [B', 'T', 'B], [2', '7', 'z', 'i', 'n', 'c], [5', '1', 'z', 'i', 'n', 'c], [B', 'C', 'L', '6', 'S], [M', 'i
l', 'y', 's', 'i', 'n', 'a', 't', 'r', 'i', 'o', 'p', 'e', 'p', 't', 'i', 'd', 'a', 's', 'e', 'c', 'o', 'm', 'm', 'o', 'n], [a', 'n,
m', 'b', 'a', 'n', 'e], [m', 'e', 't', 'a', 'l', 'l', 'o', 'e', 'n', 'd', 'o', 'p', 'e', 'p', 't', 'i', 'd', 'a', 's', 'e], [n
[e', 'n', 'd', 'o', 'p', 'e', 'p', 't', 'i', 'd', 'a', 's', 'e], [e', 'n', 'k', 'e', 'p', 'h', 'a', 'l', 'i', 'n', 'a', 's', 'e], [C
l', '0], [m', 'e', 'm', 'b', 'e', 'r', 'a', 'n', 'e], [2', 'n', 'e', 'p', 'r', 'i', 'l', 'y', 's', 'i', 'n', '3', '9', '0', 'n', 'e', 'p
n', '4', '1', '1', 'n', 'e', 'u', 't', 'r', 'a', 'l], [2', '4', '1', '1', 's', 'k', 'i', 'n], [f', 'i', 'b', 'r', 'o', 'b', 'l', 'a
t', 'a', 's', 'e], [S', 'H', '2', 'B', '3], [S', 'H', '2', 'B], [3', 'l', 'y', 'm', 'p', 'h', 'o', 'c', 'y', 't', 'e', 's', 'p,
n', 'k', 's', 'i', 'g', 'n', 'a', 'l], [L', 'n', 'k], [Z', 'A', 'P', '7', '0], [Z', 'A', 'P', '7', '0', '7', '0], [Z', 'e', 't,
a', 't', 'e', 'd], [p', 'r', 'o', 't', 'e', 'i', 'n', '7', '0], [Z', 'e', 't', 'a', 'c', 'h', 'a', 'i', 'n], [p', 'r', 'o', 't,
e', 'l', 'a', 't', 'e', 'd], [k', 'i', 'n', 'a', 's', 'e', 'z', 'e', 't', 'a], [T], [7', '0', 'k', 'D', 'a', 'z', 'e', 't', 'a,
R], [7', '0', 'k', 'D], [T', 'H', 'B', 'D], [t', 'h', 'r', 'o', 'm', 'b', 'o', 'm', 'o', 'd', 'u', 'l', 'i', 'n', 'C', 'D', '1', 'A
e', 'n', 'f', 'e', 't', 'o', 'm', 'o', 'd', 'u', 'l', 'i', 'n], [H', 'L', 'A', 'B], [B', 'H', 'L', 'A], [H', 'L', 'A', 'B', 'H,
p', 'r', 'o', 't', 'e', 'i', 'n', 'M', 'H', 'C], [a', 'n', 't', 'i', 'g', 'e', 'n', 'M', 'H', 'C], [S', 'H', 'C', 'H', 'A', 'M', 'H
u', 'l', 'e', 'l', 'e', 'u', 'k', 'o', 'c', 'y', 't', 'e], [I', 'B], [M', 'T', 'R], [m', 'e', 't', 'h', 'i', 'o', 'n', 'i', 'n,
s', 'e', '5', 'm', 'e', 't', 'h', 'y', 'l', 't', 'e', 't', 'r', 'a', 'h', 'y', 'd', 'r', 'o', 'f', 'o', 'l', 'a', 't', 'e', 'h', 'o', 'm
i', 'n', 'e], [l', 'c', 'o', 'b', 'a', 'l', 'a', 'm', 'i', 'n', 'd', 'e', 'p', 'e', 'n', 'd', 'e', 'n', 't], [s', 'y', 'n', 't', 'h
a', 'm', 'i', 'n', 'B', '1', '2], [d', 'e', 'p', 'e', 'n', 'd', 'e', 'n', 't], [I', 'G', 'F', '1], [i', 'n', 's', 'u', 'l', 'i,
n', 's', 'u', 'l', 'i', 'n', 'l', 'i', 'k', 'e], [s', 'o', 'm', 'a', 't', 'o', 'm', 'e', 'd', 'i', 'n], [I', 'B', 'm', 'e', 'c', 'h
t', 'o', 'r', 's', 'o', 'm', 'a', 't', 'o', 'm', 'e', 'd', 'i', 'n], [M', 'A', 'P', '2', 'K', '1], [d', 'u', 'a', 'l], [s,
i', 't', 'y], [I', 'E', 'R', 'K], [l', 'R', 'A', 'P', 'K', 'E', 'R', 'K], [l', 'M', 'A', 'P', 'K', 'K], [l', 'M', 'E', 'K], [I
2], [2', 'c', 'a', 'n', 'a', 'l', 'i', 'c', 'u', 'l', 'a', 'r], [p', 'r', 'o', 't', 'e', 'i', 'n', 'm', 'u', 'l', 't', 'i', 'd', 'r
[t', 'a', 'u', 'G], [b', 'e', 't', 'a', '1', 'g', 'a', 'm', 'm', 'a', '2], [s', 'u', 'b', 'u', 'n', 'i', 't', 'i', 'n', 't', 'e,
[l', 'P', 'H', 'F', 't', 'a', 'u', 'n', 'e', 'u', 'r', 'o', 'f', 'i', 'b', 'r', 'i', 'l', 'a', 'r', 'y], [t', 't', 'a', 'n', 'g', 'l
i', 'n', 'p', 'a', 'i', 'e', 'd], [h', 'e', 'l', 'i', 'c', 'a', 'l], [f', 'i', 'l', 'a', 'm', 'e', 'n', 't', 't', 'a', 'u', 'p
[l', '0', '3], [L', 'E', 'P], [l', 'e', 'p', 't', 'i', 'n', 'l', 'e', 'p', 't', 'i', 'n], [o', 'b', 'e', 's', 'i', 't', 'y], [o
o', 't', 'e', 'i', 'n', 'o', 'b', 'e', 's', 'e], [o', 'f', 'o', 'b', 'e', 's', 'i', 't', 'y], [F', 'P', 'G', 'S], [f', 'o', 'l',
l', 'u', 't', 'a', 'm', 'a', 't', 'e], [m', 'i', 't', 'o', 'c', 'h', 'o', 'n', 'i', 'd', 'r', 'i', 'a', 'l', 'f', 'o', 'l', 'y', 'l', 'p,
a', 'g', 'l', 'u', 't', 'a', 'm', 'a', 't', 'e], [s', 'y', 'n', 't', 'h', 'e', 't', 'a', 's', 'e', 't', 'e', 't', 'r', 'a', 'h', 'y,
t', 'e], [s', 'y', 'n', 't', 'h', 'a', 's', 'e', 't', 'e', 't', 'r', 'a', 'h', 'y', 'd', 'r', 'o', 'f', 'o', 'l', 'y', 'l', 'p', 'o,

28/38

'0'], ['r', 'e', 'c', 'e', 'p', 't', 'o', 'r', 'l', 'e', 'u', 'k', 'o', 'c', 'y', 't', 'e'], ['L', 'e', 'u', '1', '6', 'm', 'e', 'm', 'b', 'a', 'n', 'n', 'i', 'n', 'g'], [4, 'd', 'o', 'm', 'a', 'i', 'n', 's'], ['E', 'Z', 'H', '2'], ['E', 'Z', 'H', '2', 'e', 'n', 'h', 'a', 't', 'e'], [2, 'l', 'y', 's', 'i', 'n', 'e'], ['F', 'O', 'X', 'P', 'I'], [P, '1', 'f', 'o', 'r', 'k'], [h, 'e', 'a', 'd', 'r', 'e', 'g', 'l', 'u', 't', 'a', 'm', 'i', 'n', 'i', 'c', 'h'], [1, 'm', 'a', 'c', '1', 'r', 'e', 'g', 'u', 'l', 'a', 't', 'e', 'd'], [p, 'e', 'r', 'o', 'x', 'i', 'd', 'e'], [d, 'i', 's', 'm', 'u', 't', 'a', 's', 'e', 'e', 'p', 'i', 'd', 'i', 'd', 'y', 'm', 'i', 's'], [p, 'r', 'o', 't', 'e', 'i', 'n'], [d, 'i', 's', 'm', 'u', 't', 'a', 's', 'e', 'e', 'p', 'i', 'd', 'i', 'd', 'y', 'm', 'i', 's'], [p, 'r', 'o', 't', 'e', 'i', 'n', 'e', 'n', 'o', 'l', 'o', 'x', 'i', 'd', 'a', 's', 'e'], [B, 'm', 'a', 'n', 'g', 'a', 'n', 'e', 's', 'e', 'c', 'o', 'n', 't', 'a', 'i', 's', 'm', 'u', 't', 'a', 's', 'e', 'm', 'g', 'a', 'n', 'g', 'a', 'n', 'o', 's', 'u', 'p', 'e', 'r', 'o', 'x', 'i', 'd', 'e'], [d, 'i', 's', 'u', 'p', 'e', 'r', 'o', 'x', 'i', 'd', 'e'], [M, 'A', 'P', 'K', '3'], [3, 'M', 'A', 'P'], [p, '4', '4', 'M', 'A', 'P', 'K'], [1, 'l', 'e', 'u', 'l', 'a', 'r'], [s, 'i', 'g', 'n', 'a', 'l', 'r', 'e', 'l', 'a', 't', 'e', 'd'], [1, 'i', 'n', 's', 'u', 'l', 'i', 'n', 't', 'e', 'd'], [M, 'A', 'P', '2'], [k, 'i', 'n', 'a', 's', 'e', 'm', 'i', 'c', 'r', 'o', 'o', 't', 'u', 'b', 'u', 'l', 'e', 'a', 's', 's', 'd'], [M, 'M', 'P', '1'], [i, 'n', 't', 'e', 'r', 's', 't', 'i', 't', 'i', 'a', 'l'], [c, 'o', 'l', 'l', 'a', 'g', 'e', 'a', 'b', 'l', 'a', 's', 't'], [c, 'o', 'l', 'l', 'a', 'g', 'e', 'e', 'n', 'a', 's', 'e', 'm', 'a', 't', 'r', 'n', 'i', 'x'], [m, 'e', 't', 'a', 'e', 'a', 's', 'e'], [1, 'm', 'a', 't', 'r', 'i', 'x'], [S, 'P', '1'], [S, 'P', '1', 's', 'p', 'e', 'c', 'i', 'f', 'i', 'c', 'i', 'P', '1, 'J', 'u', 'n'], [p, 'r', 'o', 't', 'e', 'i', 'n', 'a', 'c', 't', 'i', 'v', 'a', 't', 'o', 'n'], [1, 'e', 'n', 'h', 'a', 'n', 'd', 'i', 'n', 'g'], [A, 'P', '1, 'j', 'u', 'n'], [o', 'n', 'c', 'o', 'g', 'e', 'n', 'e', 'p', 'r', 'o', 't', 'o', 'n', 'c', 'o', 'u', 'n', 'v', 'j', 'u', 'n'], [h, 'o', 'm', 'o', 'l', 'o', 'g', 'v', 'j', 'u', 'n'], [L, 'G', 'A', 'L', 'S', '3], [g, 'a', 'l', '5], [1, 'e', 'c', 't', 'i', 'n', 'i', 'g', 'E', 'b', 'i', 'n', 'g', 'd', 'i', 'n', 'g'], [p, 'r', 'o', 't', 'e', 'i', 'n', 'M', 'A', 'C', 'e', 'n', 'a', 'd', 'v', 'a', 'n', 'c', 'e', 'd'], [e', 'n', 'd', 'p', 'r', 'o', 'd', 'u', 'c', 't], [3, 'c', 'a', 'n', 'b', 'o', 'h', 'b', 'i', 'n', 'd', 'i', 'n', 'g'], [3, '5', 'e', 'p', 'i', 'd', 'i', 'd', 'y', 'm', 'i', 's'], [p, 'r', 'o', 't', 'e', 'i', 'n', 'i', 'n', 'g', 'e', 's', 'p', 'e', 'c', 'i', 'f', 'i', 'c], [1, 'e', 'c', 't', 'i', 'n', 'i', 'g', 'a', 'c', 't', 'i', 'n', 'i', 'g], [3, '1', 'a', 'm', 'i', 'n', 'b', 'i', 'n', 'o', 's', 't', 'e', 'i', 'n', 'l', 'e', 'c', 't', 'i', 'n', 'i', 'g], [L, '2', '9', 'l', 'e', 'c', 't', 'i', 'n', 'i', 'g', 'a', 'l', 'a', 'b', 't', 'o', 'n', 'd', 'i', 'n', 'g], [P, 'T', 'K', '2], [f, 'o', 'c', 'a', 'l], [1, 'F', 'A', 'D', 'K], [1, 'F', 'A', 'K', 'r', 'e', 'l', 'n', 'k', 'i', 'n', 'a', 's', 'e], [p, 'o', 'l', 'y', 'p', 'e', 'p', 't', 'i', 'd', 'e', 'P', 'T', 'K', '2], [2, 'f', 'o', 'c', 'a', 'l', '3, '3', 'f', 'o', 'c', 'a', 'l], [k, 'i', 'n', 'a', 's', 'e', 'r', 'e', 'l', 'a', 't', 'e', 'd], [n', 'o', 'n', 'k', 'i', 'n', 't', 'e', 'i', 'n', 'i', '7', '1], [M, 'E', 'F', '2, 'C], [m, 'y', 'o', 'c', 'y', 't', 'e', 's', 'p', 'e', 'c', 'i', 'f', 'i', 'c], [M, 'E', 'R', 'T', 'K], [M, 'E', 'r', 'M', 'E', 'R], [k, 'i', 'n', 'a', 's', 'e', 'S', 'T', 'K], [k, 'i', 'n', 'a', 's', 'e', 'e', 'r', 'n', 'e', 'c', 'e', 'p', 't', 'o', 'r', 'n], [M, 'E', 'r', 'T', 'K], [B, 'I', 'R', 'C', '7], [7, 'R', 'I', 'N', 'G], [S, 'p', 'e', 'l], [B, 'I', 'R', 'C', '7], [k, 'i', 'd', 'n', 'e', 'y], [p, 'r', 'o', 't', 'e', 'i', 'n', 'l', 'i', 'v', 'i', 'n], [a, 's', 'm', 'e', 'l', 'a', 'n', 'o', 'm', 'a], [H, 'L', 'A', 'C], [C, 'w', 'l], [c, 'h', 'a', 'i', 'n', 'H', 'L', 'A], [c, 'h', 'C], [H, 'L', 'A', 'C, 'h', 'u', 'm', 'a', 'n], [a', 'n', 't', 'i', 'g', 'e', 'n', 'C], [c, 'h', 'a', 'i', 'n', 'm', 'a', 'j', 'i', 'g', 'a', 't', 'i', 'o', 'n], [f, 'i', 'a', 'c', 't', 'o', 'r', 'L', 'd', 'o', 'p', 'a', 'c', 'h', 'i', 'r', 'o', 'm', 'e', 'i', 's', 'e', 'L', 'd', 'o', 'p', 'a', 'c', 'h', 'i', 'r', 'o', 'm', 'e', 'i], [t, 'a', 'u', 't', 'o', 'm', 'e', 'r', 'a', 's', 'e', 'e', 'p', 'i', 'd', 'p, 'r', 'o', 't', 'e', 'i', 'n', 'm', 'a', 'c', 'r', 'o', 'p', 'h', 'a', 'g', 'e], [g, 'l', 'y', 'c', 'o', 's', 'y', 'l', 'a', 't', 'b', 'i', 't', 'i', 'n', 'g], [p, 'h', 'e', 'n', 'l', 'p', 'y', 'h', 'u', 'v', 'a', 't', 'e], [t, 'a', 'u', 't', 'o', 'm', 'e', 'W', 'R], [P, 'R', 'K', 'C], [p, 'r', 'o', 't', 'e', 'i', 'n', 'P', 'R', 'K', 'C], [r', 'e', 'g', 'u', 'l', 'a', 't', 'o', 'r', 'r', 'a', 'c', 't', 'i', 'n', 'g], [p, 'r', 'o', 't', 'e', 'i', 'n', 'p', 'r', 'o', 's', 't', 'a', 't', 'e], [4, 'p', 'r', 'o', 's', 'R', '4], [D, 'L', '4], [d, 'e', 'l', 't', 'a', 'l', 'i', 'k', 'e], [4, 'd', 'e', 'l', 't', 'a', 'e], [h, 'o', 'm', 'o', 'l', 'o', 'g', 'd', 'e', 'l', 't', 'a', 'l', 'i', 'k', 'e], [p, 'r', 'o', 't', 'e', 'i', 'n', 'd', 'r', 'o', 'a], [D, 'e', 'l', 't', 'a], [4, 'n', 'o', 't', 'c', 'h], [d, 'e', 'l', 't', 'a], [d, 'e', 'l', 't', 'a], [D, '1, '9', 'B', '1, 'y', 'm', 'p', 'h', 'o', 'c', 'y', 't', 'e], [B, '4', 'T', 'c', 'e', 'l', 'l], [L, 'e', 'u', '1, '2', 'd', 'n', 't', 'i', 'a', 't', 'i', 'o', 'n], [K, 'C', 'N', 'H', '2], [2, 'e', 'a', 'g], [h, 'o', 'm', 'o', 'l', 'o', 'g', 'e', 'a', 'd', 'l], [1, 'e', 't', 'h', 'e', 'r', 'a', 'g', 'o', 'g', 'o', 'r', 'e', 'l', 'a', 't', 'e', 'd], [p, 'r', 'o', 't', 'e', 'i', 'n', 'e', 'o', 'g', 'o', 'r', 'e', 'l', 'a', 't', 'e', 'd], [1, 'p', 'o', 't', 'a', 's', 's', 'i', 'u', 'm], [e', 'a', 'g], [2, 'v', 'o', 'l', 't', 'a', 'g', 'e', 'g', 'a', 't', 'e', 'd], [K, 'v', '1, '1, '1], [S, 'T', 'A', 'T', '1], [1, 'a', 'l', 's', 'i', 'g', 'n', 'a', 'l], [9, 'l', 'k', 'D', 's', 'i', 'g', 'n', 'a', 'l], [9, 'l', 'k', 'D', 'a', 't', 'r', 'n', 'a', 'n', 's', 'c', 'n], [I, 'S', 'G', 'F', '3], [c, 'o', 'm', 'p', 'o', 'n', 'e', 'n', 't', 's], [p, '9, '1, 'p', '8, '4], [I, 'D', 'H', '1', 'r', 'a', 'a', 't', 'e], [N, 'A', 'D', 'A], [c, 'a', 'y', 't', 'o', 'p', 'l', 'a', 's', 'm', 'i', 'c', 'N', 'A', 'D', 'P], [I, 'C', 'D', 'p', 'e', 'n', 'd', 'e', 'n', 't], [c, 'y', 't', 'o', 's', 'o', 'l', 'i', 'c', 'N', 'A', 'D', 'P], [d, 'e', 'p', 'e', 'n', 'd', 'e', 'x', 'i', 's', 'o', 'm', 'a', 'l', 'e', 'p', 'i', 'd', 'i', 'd', 'y', 'm', 'i', 's], [1, 'u', 'm', 'i', 'n', 'a', 'l], [2, '1, '6', 'y', 'i', 's], [2, '6', 'e', 'p', 'i', 'd', 'i', 'd', 'y', 'm', 'i', 's], [p, 'r', 'o', 't', 'e', 'i', 'n', 'i', 's', 'o', 'c', 'N', 'A', 'D', 'P', '+], [s', 'o', 'l', 'u', 'b', 'l', 'e', 'o', 'x', 'a', 'l', 'o', 's', 'u', 'c', 'c', 'i', 'n', 'a', 't', 'e], [o', 'm', 'b', 'i', 'n', 'a', 't', 'i', 'o', 'n', 'a', 'c', 't', 'i', 'v', 'a', 't', 'i', 'n', 'g], [1, 'R', 'I', 'N', 'G], [7, '4', 'i', 'n', 'a', 't', 'i', 'o', 'n], [1, 'r', 'e', 'c', 'o', 'm', 'b', 'i', 'n', 'a', 't', 'i', 'o', 'n], [8, 'F', 'A', 'D', 'h', 'o', 'm', 'o', 'l', 'o', 'g', 'o', 'u', 's], [I, 'C', 'E', 'C', 'E', 'D', '3, '1, 'i', 'k', 'e], [p, 'e', 'F', 'A', 'D', 'D', 'l', 'i', 'k', 'e], [I, 'C', 'E', 'I', 'C', 'E', 'l', 'i', 'k', 'e], [a', 'p', 'o', 'p', 't', 'o', 't', 'i', 'a', 's', 'e], [5, 'M', 'A', 'C', 'H', 'a', 'l', 'p', 'h', 'a', '1, '2, '3], [p, 'r', 'o', 't', 'e', 'i', 'n', 'M', 'A', 'C', 'H', '3, '4], [p, 'r', 'o', 't', 'e', 'i', 'n', 'M', 'O', 'R', 'A', 'T], [1, 'a', 's', 's', 'i', 'o', 'c', 'i', 'a', 't', 'e', 'd], [c, 'e', 'h', 'o', 'g', 'a', 'p', 'o', 'p', 't', 'o', 't', 'i', 'c], [p, 'r', 'o', 't', 'e', 'i', 'a', 's', 'e', 'a', 'p', 'o', 'p', 't', 'o', 't', 'i', 'a', 's', 'e], [a', 'p', 'o', 'p', 't', 'o', 't', 'i', 's', 'p', 'a', 's', 'e], [a', 'p', 'o', 'p', 't', 'o', 't', 'i', 's', 'p', 'a', 's', 'e], [2, '6', 'e', 'p', 'i', 'd', 'i', 'd', 'y', 'm', 'i', 's], [p, 'r', 'o', 't', 'e', 'i', 'n', 'i', 's', 'o', 'c', 'N', 'A', 'D', 'P', '+], [s', 'o', 'l', 'u', 'b', 'l', 'e', 'o', 'x', 'a', 'l', 'o', 's', 'u', 'c', 'c', 'i', 'n', 'a', 't', 'e], [o', 'm', 'b', 'i', 'n', 'a', 't', 'i', 'o', 'n', 'a', 'c', 't', 'i', 'v', 'a', 't', 'i', 'n', 'g], [1, 'R', 'I', 'N', 'G], [7, '4', 'i', 'n', 'a', 't', 'i', 'o', 'n], [1, 'r', 'e', 'c', 'o', 'm', 'b', 'i', 'n', 'a', 't', 'i', 'o', 'n], [8, 'F', 'A', 'D', 'h', 'o', 'm', 'o', 'l', 'o', 'g', 'o', 'u', 's], [I, 'C', 'E', 'C', 'E', 'D', '3, '1, 'i', 'k', 'e], [p, 'e', 'F', 'A', 'D', 'D', 'l', 'i', 'k', 'e], [I, 'C', 'E', 'I', 'C', 'E', 'l', 'i', 'k', 'e], [a', 'p', 'o', 'p', 't', 'o', 't', 'i', 'a', 's', 'e], [5, 'M', 'A', 'C', 'H', 'a', 'l', 'p', 'h', 'a', '1, '2, '3], [p, 'r', 'o', 't', 'e', 'i', 'n', 'M', 'A', 'C', 'H', '3, '4], [p, 'r', 'o', 't', 'e', 'i', 'n', 'M', 'O', 'R', 'A', 'T], [1, 'a', 's', 's', 'i', 'o', 'c', 'i', 'a', 't', 'e', 'd], [c, 'e', 'h', 'o', 'g', 'a', 'p', 'o', 'p', 't', 'o', 't', 'i', 'c], [p, 'r', 'o', 't', 'e', 'i', 'a', 's', 'e', 'a', 'p', 'o', 'p', 't', 'o', 't', 'i', 's', 'p', 'a', 's', 'e], [a', 'p', 'o', 'p', 't', 'o', 't', 'i', 's', 'p', 'a', 's', 'e], [2, '6', 'e', 'p', 'i', 'd', 'i', 'd', 'y', 'm', 'i', 's], [p, 'r', 'o', 't', 'e', 'i', 'n', 'i', 's', 'o', 'c', 'N', 'A', 'D', 'P', '+], [s', 'o', 'l', 'u', 'b', 'l', 'e', 'o', 'x', 'a', 'l', 'o', 's', 'u', 'c', 'c', 'i', 'n', 'a', 't', 'e], [h, 's', 'a', 'm', 'i', 'r', '1, '4, '6, 'a], [F, 'L', 'T', '1], [1, 'f', 'm', 's', 'l', 'i', 'k', 'e], [1, 'f', 'm', 's', 'l', 'f', 'a', 'c', 't', 'o', 'r', 'v', 'a', 's', 'c', 'u', 'l', 'a', 'r], [F, 'R', 'T', 't', 'y', 'r', 'o', 's', 'i', 'n', 'e', 'p', 'r', 'l', 'L', 'T', 'v', 'a', 's', 'c', 'u', 'l', 'a', 'r], [M, 'I', 'R', '1, '4, '2], [h, 's', 'a', 'm', 'i', 'r', '1, '4, '2], [H, 'c', 'e', 't', 'y', 'l', 'a', 's', 'e], [d, 'e', 'p', 'e', 'n', 'd', 'e', 'n', 'c', 'y], [y', 'e', 'a', 's', 't], [h, 'o', 'm', 'e', 'e], [O, 'P', 'R', 'M', '1], [m', 'u', 't', 'y', 'p', 'e], [o', 'p', 'i', 'o', 'i', 'd], [r', 'e', 'c', 'e', 'p', 't', 'o', 'r', 't', 'e], [h, 'M', 'O', 'R', '1, 'a], [S, 'L', 'C', '2, 'A', '1], [f, 'a', 'c', 'i', 'l', 'i', 't', 'a', 't', 'e', 'd], [g, '1, 'c', 'h', 'o', 'r', 'e', 'o', 'a', 't', 'h', 'e', 't', 'o', 's', 'i', 's', 'p', 'a', 's', 't', 'i', 'c], [i', 'c', 'y], [e', 'c], [p, 'a', 'r', 'o', 'x', 'y', 's', 'm', 'a', 'l], [c, 'h', 'o', 'r', 'e', 'o', 'a', 't', 'h', 'e', 't', 'o', 's', 'i', 's', 'i', 't', 'y], [e', 'r', 'y], [e', 'r', 'y], [h, 'r', 'y], [t', 'h', 'r', 'o', 'c', 'y], [t', 'e', 'b', 'r', 'n', 'a', 'i', 'n', 'h', 'e', 'p', 'G', '2], [t', 'r', 'a', 'e', 'r', 'h', 'u', 'm', 'a', 'n], [r', 'e', 'p', 't', 'o', 'r', 'r', 'e', 'c', 'e', 'p', 't', 'o', 'r], [H, 'T', 'L', 'V', 's', 'o', 'l', 'u', 't', 'e], [T, 'N', 'F', 'R', 'S', 'F', '1, 'A], [1, 'A', 'T', 'N', 'F', 'R', '1, 'T', 'N', 'F', 'R', 'I', 'T', 'm', 'o', 'r', 'l], [1, 't', 'u', 'm', 'o', 'r], [b', 'e', 't', 'a', 't', 'u', 'm', 'o', 'r], [H, 'L', 'F], [h, 'e', 'p', 'a', 't', 'o', 'r', 'H, 'L', 'F], [P, 'A', 'R], [b', 'Z', 'I', 'P], [f, 'a', 'c', 't', 'o', 'r', 'P, 'A', 'R], [T, 'P', '7, '3], [i', 'k', 'e], [f, 'a', 'c', 't', 'o', 'r], [p', 'o', 'r', '5, '3, 'n', 'e', 'l', 'a', 't', 'e], [F, 'A', 'S', 'L', 'G], [6, 'C', 'n', 'd', 'F', 'a', 's], [A, 'P', 'O', '1], [l', 'i', 'g', 'a', 'n', 'd', 'f', 'a', 's], [l', 'i', 'g', 'a', 'n', 'd', 'm', 'u', '2, 'E', '1], [2, 'E', '1, '4, 'n', 'i', 't', 'r', 'o', 'p', 'h', 'e', 'n', 'o', 'l], [2, 'h', 'y', 'd', 'r', 'o', 'x', 'y', '1, 'I', 'E', '1, 'c', 'y', 't', 'o', 'c', 'h', 'h', 'r', 'o', 'm', 'e], [I, 'I', 'E], [e', 't', 'h', 'a', 'n', 'o', 'l', 'i', 'n', 'd', 'P, '4, '5, 'J', 'f', 'l', 'a', 'v', 'o', 'p', 'r', 'o', 't', 'e', 'i', 'n', 'l', 'i', 'n', 'k', 'e], [I, 'G', 'F', '2, 'A', 'b', 'i', 'n', 'd', 'i', 'n', 'g], [1, 'I', 'G', 'F', 'I, 'I], [1, 'I', 'G', 'F', '2], [1, 'V', 'I', 'C', 'K', 'Z], [i', 'n', 'g], [d, 'e', 't', 'e', 'r', 'm', 'i', 'n', 'a', 'n', 't], [p, 'r', 'o', 't', 'e', 'i', 'n', 'l', 'i', 'n', 'k', 'e], [m, 'R', 'N', 'A], [d, 'e', 'l', 't', 'a', 'n], [C, 'R', 'D', 'B', 'Z', 'i', 'p', 'c', 'o', 'd', 'e', 'g], [C, 'D', '2, '7], [a', 'n', 't', 'i', 'g', 'e', 'n', 'C', 'D', '2, '7, 'L], [r', 'e', 'c', 'e', 'p', 't', 'o', 'r', 'T], [1, 'l], [C, 'D', '2, '7, 't', 'u', 'm', 'o', 'r], [P, 'L', 'K', '1], [P, 'L', 'K', '1, 'P, 'L', 'K', '1, 'c', 'e', 'l', 't', 'e', 'd], [k, 'i', 'n', 'a', 's', 'e', 'p', 'o', 'l', 'o', 'l], [D, 'r', 'o', 's', 'o', 'p', 'h', 'i', 'i', 'n', 'a', 'e', 't', 'h', 'r', 'e', 'e', 'p', 'r', 'o', 't', 'e', 'i', 'n], [1, '3], [S, 'M', 'A', 'D', '3], [m, 'o', 'c', 'a', 'p', 'e', 'n', 't', 'a', 'p', 'l', 'e', 'g', 'i', 'c], [3, 'M', 'A', 'D], [3, 'S', 'M', 'A], [M, 'A', 'D', 'r', 'e', 'M, 'A', 'D], [D, 'P], [3, 'h', 'M, 'A', 'D, '3, 'h', 'S, 'M, 'A', 'D, '3, 'm', 'a', 'a', 'V, '1, '5, '2, '1, 'o', 'g', 'm', 'a', 'd', '3, 'm', 'o', 't', 'h', 'e', 'r', 's], [C, 'C', 'N], [2], [c, 'o', 'n', 'n', 'e', 'c', 't', 'i', 'v', 'e], [f, 'a', 'c', 't', 'o', 'r], [C, 'C', 'N], [2, 'I', 'B', 'P, '8, 'I', 'G', 'F', 'b', 'i', 'n', 'd', 'i', 'n', 'g], [8, 'y', 'p', 'e', 'r', 't', 'r', 'o', 'p', 'h', 'i', 'c], [c, 'h', 'o', 'n', 'd', 'r', 'o', 'c', 'y', 't', 'e', 's', 'p', 'e', 'c', 'i,

'n', 's', 'u', 'l', 'i', 'n', 'l', 'i', 'k', 'e', ['f', 'a', 'c', 't', 'o', 'n', 'b', 'i', 'n', 'd', 'i', 'n', 'g'], ['M', 'I', 'R', '1',
'i', 'r', '1', '0', '0'], ['C', 'S', 'N', 'K', '2', 'A', '1'], ['c', 'a', 's', 'e', 'i', 'n'], ['a', 'l', 'p', 'h', 'a', 'C', 'K', '2'],
's', 'e', 'i', 'n'], ['p', 'o', 'l', 'y', 'p', 'e', 'p', 't', 'i', 'd', 'e', 'c', 'a', 's', 'e', 'i', 'n'], ['p', 's', 'e', 'u', 'd', 'o',
'e', 'i', 'n'], ['s', 'u', 'b', 'u', 'n', 'i', 't', 'p', 'r', 'o', 't', 'e', 'i', 'n'], ['C', 'K', '2'], ['D', 'P', 'P', '4'], ['d', 'i',
'l'], ['4', 'A', 'D', 'C', 'P', '2', 'D', 'P', 'P'], ['I', 'V', 'T', 'c', 'e', 'l', 'l'], ['C', 'K', '2', '6', 'a', 'd', 'e', 'n', 'o',
'm', 'i', 'n', 'a', 's', 'e'], ['c', 'o', 'm', 'p', 'l', 'e', 'x', 'i', 'n', 'g'], ['2', 'd', 'i', 'p', 'e', 'p', 't', 'i', 'd', 'y', 'l',
'p', 't', 'i', 'd', 'y', 'l', 'p', 'e', 'p', 't', 'i', 'd', 'a', 's', 'e'], ['4', 'd', 'i', 'p', 'e', 'p', 't', 'i', 'd', 'y', 'l', 'p',
'e'], ['C', 'D', '2', '6'], ['C', 'A', 'L', 'C', 'A'], ['c', 'a', 'l', 'c', 'i', 't', 'o', 'n', 'i', 'n', 'c', 'a', 'l', 'c', 'i', 't',
'e', 'r', 'e', 'l', 'a', 't', 'e', 'd'], ['1', 'a', 'l', 'p', 'h', 'a', 't', 'y', 'p', 'e'], ['C', 'G', 'R', 'P', 'c', 'a', 'l', 'c', 'i',
'c', 'a', 'l', 'c', 'i', 't', 'o', 'n', 'i', 'n'], ['I', 'c', 'a', 'l', 'c', 'i', 't', 'o', 'n', 'i', 'n', 'c', 'a', 'l', 'c', 'i', 't',
'a', 't', 'e', 'd'], ['a', 'l', 'p', 'h', 'a', 'k', 'a', 't', 'a', 'c', 'a', 'l', 'c', 'i', 'n'], ['Z', 'N', 'F', '4', '2', '3'], ['4',
'B', 'F'], ['f', 'i', 'n', 'g', 'e', 'r', 'S', 'm', 'a', 'd'], ['O', 'l', 'f', 'i', 'n', 't', 'e', 'r', 'a', 'c', 't', 'i', 'n', 'g'], [
'e', 'a', 'r', 'l', 'y'], ['p', 'r', 'o', 't', 'e', 'i', 'n', 'o', 'l', 'f', '1', 'E', 'B', 'F', 'a', 's', 's', 'o', 'c', 'i', 'a', 't',
'l'], ['O', 'l', 'f', 'o', 'r', 'K', 'h', 'e', 'a', 'd'], ['O', '1', 'A', 'f', 'o', 'r', 'k', 'h', 'e', 'a', 'd'], ['r', 'h', 'a', 'b',
'r', 'c', 'o', 'm', 'a'], ['X', 'I', 'A', 'P'], ['X', 'I', 'A', 'P', '1', 'i', 'k', 'e'], ['p', 'r', 'o', 't', 'e', 'i',
'e'], ['1', 'R', 'I', 'N', 'G', 't', 'y', 'p', 'e'], ['X', 'I', 'A', 'P', 'X', 'l', 'i', 'n', 'k', 'e', 'd'], ['I', 'A', 'P', 'X', 'l',
'g', 'a', 's', 'e', 'b', 'a', 'c', 'u', 'l', 'o', 'v', 'i', 'r', 'a', 'l'], ['4', 'i', 'n', 'h', 'i', 'b', 'i', 't', 'o', 'r'], ['I', 'R',
'l', 'i', 'n'], ['s', 'u', 'b', 's', 't', 'r', 'a', 't', 'e'], ['1', 'I', 'R', 'S', '1'], ['D', 'U', 'X', '4'], ['4', 'd', 'o', 'u', 'b',
'e', 'd', 'o', 'u', 'b', 'l', 'e'], ['1', '0', 'd', 'o', 'u', 'b', 'l', 'e'], ['4', '1', '0'], ['P', 'D', 'G', 'F', 'R', 'A'], ['p', 'l',
'e', 'r', 'i', 'v', 'e', 'd'], ['a', 'l', 'p', 'h', 'a', 'C', 'D', '1', '4', '0'], ['A', 'C', 'D', '1', '4', '0', 'a'], ['a', 'n', 't', 'i',
'F', 'R', 'a', 'l', 'p', 'h', 'a', 'l', 'p', 'h', 'a', 't', 'y', 'p', 'e'], ['r', 'e', 'c', 'e', 'p', 't', 'o', 'r', 'p', 'l', 'a',
'r', 'i', 'v', 'e', 'd'], ['2', 'p', 'l', 'a', 't', 'e', 'l', 'e', 't', 'd', 'e', 'r', 'i', 'v', 'e', 'd'], ['F', 'O', 'X', 'M', '1'], [
'e', 'a', 'd'], ['d', 'r', 'o', 's', 'o', 'p', 'h', 'i', 'l', 'a'], ['1', '6', 'H', 'N', 'F', '3', 'f', 'o', 'r', 'k', 'h', 'e', 'a', 'd
's', 'e'], ['p', 'h', 'o', 's', 'p', 'h', 'o', 'p', 'r', 'o', 't', 'e', 'i', 'n'], ['2', 'M', 'P', 'M', '2'], ['r', 'e', 'a', 'c', 't',
'l'], ['F', 'O', 'X', 'M', '1', 'D', 'f', 'o', 'r', 'k', 'h', 'e', 'a', 'd'], ['M', '1', 'D', 'f', 'o', 'r', 'k', 'h', 'e', 'a', 'd', 'r',
'F', 'K', 'H', 'L', '1', '6', 'h', 'e', 'p', 'a', 't', 'o', 'c', 'y', 't', 'e'], ['1', 'l', 't', 'r', 'a', 'n', 's', 'c', 'r', 'i', 'p',
'i', 'd', 'e', 'n', 't', 'w', 'i', 'n', 'g', 'e', 'd', 'h', 'e', 'l', 'i', 'x'], ['I', 'N', 'S', '1'], ['D', 'K', 'K', '1'], ['d', 'i',
'e', 'l', 'a', 't', 'e', 'd'], ['1', 'd', 'i', 'c', 'k', 'k', 'o', 'p', 'f'], ['h', 'o', 'm', 'o', 'l', 'o', 'g', 'd', 'i', 'c', 'k', 'k',
'k', 'e', 'd', 'i', 'c', 'k', 'k', 'o', 'p', 'f', 'l', 'i', 'k', 'e'], ['P', 'R', 'K', 'A', 'C', 'A'], ['c', 'A', 'M', 'P', 'd', 'e', 'p',
'a', 'l', 'p', 'h', 'a', 'P', 'K', 'A'], ['C', 'a', 'l', 'p', 'h', 'a', 'p', 'r', 'o', 't', 'e', 'i', 'n'], ['H', 'S', 'P', 'A', 'l', 'i',
'h', 'o', 'c', 'k'], ['7', '0'], ['1', 'A', 'H', 'S', 'P', '7', '0', '1', 'H', 'S', 'P', '7', '0', '2', 'H', 'S', 'P', '7', '0', '1', 'H
'e', 'a', 't'], ['1', 'B', 'H', 'e', 'a', 't'], ['2', 'd', 'i', 'n', 'a', 'K', 't', 'y', 'p', 'e'], ['m', 'o', 'l', 'e', 'c', 'u', 'l', 'a',
'o', 'n', 'e'], ['H', 'S', 'P', '7', '0', '1', 'e', 'p', 'i', 'd', 'i', 'd', 'y', 'm', 'i', 's'], ['1', '0', '3', 'e', 'p', 'i', 'd', 'i',
'r', 'o', 't', 'e', 'i', 'n', 'h', 'e', 'a', 't'], ['1', 'h', 'e', 'a', 't'], ['1', '2', 'h', 'e', 'a', 't'], ['1', '1', 'B', 'h',
'a', 't'], ['7', '0', 'k', 'D', 'a'], ['s', 'h', 'o', 'c', 'i', 'i', 'n', 'd', 'u', 'c', 'e', 'd'], ['M', 'P', '7'], ['m', 'a', 't',
'm', 'a', 't', 'r', 'i', 'n', 'm', 'a', 't', 'r', 'i', 'x'], ['m', 'a', 't', 'r', 'i', 'l', 'y', 's', 'i', 'n'], ['u', 't', 'e', 'r', 'i
'l', 'l', 'o', 'p', 'r', 'o', 't', 'e', 'i', 'n', 'a', 's', 'e', '7', 'p', 'u', 'm', 'p', 'l'], ['p', 'r', 'o', 't', 'e', 'i', 'a', 's', 'e,
'm', 'a', 't', 'r', 'i', 'l', 'y', 's', 'i', 'n', 'u', 't', 'e', 'r', 'i', 'n', 'e'], ['S', '1', '0', '0', 'A', '9'], ['S', '1', '0', '0
'4', 'c', 'a', 'l', 'g', 'r', 'a', 'n', 'u', 'l', 'i', 'n'], ['B', 'c', 'a', 'l', 'p', 'r', 'o', 't', 'e', 'c', 't', 'i', 'n'], ['L', '1
'i', 't', 'l', 'e', 'u', 'k', 'o', 'c', 'y', 't', 'e'], ['1', '4'], ['P', 'D', 'G', 'F', 'R', 'B'], ['b', 'e', 't', 'a', 'A', 'c', 't',
'D', 'G', 'F', 'R', 'B', 'C', 'D', '1', '4', '0'], ['B', 'N', 'D', 'E', 'L', '1', 'P', 'D', 'G', 'F', 'R', 'B', 'P', 'D', 'G', 'F', 'R',
'F', 'R', 'b', 'e', 't', 'a', 'b', 'e', 't', 'a', 't', 'y', 'p', 'e'], ['1', 'p', 'l', 'a', 't', 'e', 'l', 'e', 't', 'd', 'e', 'r', 'i',
'l'], ['1', 'E', 'G', 'R', '1', 'n', 'e', 'r', 'v', 'e'], ['A', 't', 'r', 'a', 'n', 's', 'c', 'r', 'i', 'p', 't', 'i', 'o', 'n'], ['E',
'a', 'n', 's', 'c', 'r', 'i', 'p', 't', 'i', 'o', 'n'], ['Z', 'i', 'f', '2', '6', '8', 'z', 'i', 'n', 'c'], ['2', '2', 'S', 'z', 'i', 'n
'4'], ['P', 'I', 'N', '1'], ['p', 'e', 'p', 't', 'i', 'd', 'y', 'l', 'p', 'r', 'o', 'l', 'y', 'l'], ['c', 'i', 's', 't', 'r', 'a', 'n',
'a', 's', 'e'], ['N', 'I', 'M', 'A', 'i', 'n', 't', 'e', 'r', 'a', 'c', 't', 'i', 'n', 'g'], ['1', 'P', 'P', 'I', 'a', 's', 'e'], ['P', 'i
'e', 'i', 'n'], ['l', 'r', 'o', 't', 'a', 'm', 'a', 's', 'e'], ['P', 'i', 'n', 'l'], ['P', 'K', 'M'], ['p', 'y', 'n', 'u', 'v', 'a', 't',
'P', '3', 'O', 'P', 'A', 'i', 'n', 't', 'e', 'r', 'a', 'c', 't', 'i', 'n', 'g'], ['3', 'P', 'K'], ['t', 'y', 'p', 'e', 'c', 'y', 't', 'o
'h', 'y', 'r', 'o', 'i', 'd'], ['h', 'o', 'r', 'm', 'o', 'n', 'e', 'b', 'i', 'n', 'd', 'i', 'n', 'g'], ['p', 'r', 'o', 't', 'e', 'i', 'n
'y', 'm', 'i', 's'], ['3', '0', 'p', '5', '8', 'p', 'y', 'n', 'u', 'v', 'a', 't', 'e'], ['2', '3', 'p', 'y', 'r', 'u', 'v', 'a', 't', 'e
'e', 's'], ['M', '1', 'M', '2', 'p', 'y', 'r', 'u', 'v', 'a', 't', 'e'], ['i', 's', 'o', 'z', 'y', 'm', 'e', 'p', 'y', 'r', 'u', 'v', 'a
'l', 'e', 't', 'h', 'y', 'r', 'o', 'i', 'd'], ['1', 't', 'h', 'y', 'r', 'o', 'i', 'd'], ['c', 'y', 't', 'o', 's', 'o', 'i', 'c', 't',
'P', 'K'], ['F', 'H', 'I', 'T'], ['b', 'i', 's'], ['5', 'a', 'd', 'e', 'n', 'o', 's', 'y', 'l'], ['t', 'r', 'i', 'p', 'h', 'o', 's', 'p',
'P', '3', 'A'], ['h', 'y', 'd', 'r', 'o', 'l', 'a', 's', 'e', 'd', 'i', 'a', 'd', 'e', 'n', 'o', 's', 'i', 'n', 'e'], ['5', '5'], ['P',
'h', 'o', 's', 'p', 'h', 'a', 't', 'e'], ['h', 'y', 'd', 'r', 'o', 'l', 'a', 's', 'e', 'd', 'i', 'n', 'u', 'c', 'l', 'e', 'o', 's', 'i',
'o', 's', 'p', 'h', 'a', 't', 'a', 's', 'e'], ['L', 'C', 'K'], ['L', 'c', 'k', 'T', 'l', 'y', 'm', 'p', 'h', 'o', 'c', 'y', 't', 'e], [
'e', 'u', 'k', 'o', 'c', 'y', 't', 'e'], ['S', 'r', 'c'], ['k', 'i', 'n', 'a', 's', 'e', 'l', 'y', 'm', 'p', 'h', 'o', 'c', 'y', 't', 'e
'e', 'c', 'i', 'f', 'i', 'c'], ['p', 'r', 'o', 't', 'e', 'i', 'n', 't', 'y', 'r', 'o', 's', 'i', 'n', 'e'], ['k', 'i', 'n', 'a', 's', 'e
'R', 'A'], ['L', 'C', 'K', 'T'], ['B', 'M', 'P', '4'], ['b', 'o', 'n', 'n', 'e'], ['m', 'o', 'n', 'p', 'h', 'o', 'g', 'e', 'n', 'e', 't', 'i
'P', 'R', 'K', 'A', 'A', '1'], ['5', 'A', 'M', 'P', 'a', 'c', 't', 'i', 'v', 'a', 't', 'e', 'd'], ['a', 'l', 'p', 'h', 'a', '1', '5',
'v', 'a', 't', 'e', 'd'], ['a', 'l', 'p', 'h', 'a', '1'], ['c', 'h', 'a', 'i', 'n', 'A', 'C', 'A', 'C', 'A'], ['k', 'i', 'n', 'a', 's',
'i', 'v', 'a', 't', 'e'], ['s', 'u', 'b', 'u', 'n', 'i', 't', 'A', 'M', 'P', 'a', 'c', 't', 'i', 'v', 'a', 't', 'e', 'd'], ['a', 'l', 'p
'K'], ['1', 'A', 'M', 'P', 'K'], ['a', 'l', 'p', 'h', 'a', '1', 'H', 'M', 'G', 'C', 'R'], ['k', 'i', 'n', 'a', 's', 'e', 'a', 'c', 'e,
'a', 'r', 'b', 'o', 'x', 'y', 'l', 'a', 's', 'e], ['k', 'i', 'n', 'a', 's', 'e', 'h', 'y', 'd', 'r', 'o', 'x', 'y', 'm', 'e', 't', 'h',
'r', 'y', 'l', 'C', 'o', 'A'], ['k', 'i', 'n', 'a', 's', 'e', 'p', 'r', 'o', 't', 'e', 'i', 'n'], ['A', 'M', 'P', 'a', 'c', 't', 'i', 'v
'b', 'u', 'n', 'i', 't', 'a', 'u', 'p', 'r', 'o', 't', 'e', 'i', 'n'], ['S', 'M', 'A', 'D', '2'], ['2', 'M', 'A', 'D'], ['2', 'S',
'a'], ['M', 'a', 'd', 'r', 'e', 'l', 'a', 't', 'e', 'd'], ['2', 'm', 'o', 't', 'h', 'e', 'r'], ['A', 'B', 'C', 'C', '1'], ['A', 'B', 'C
'e', 'x', '1', '3', 'A', 'T', 'P', 'b', 'i', 'n', 'd', 'i', 'n', 'g'], ['A', 'B', 'C', 'C', '1', 'd', 'e', 'l', 't', 'a', 'e', 'x', '1,
'i', 'n', 'd', 'i', 'n', 'g'], ['A', 'B', 'C', 'C', '1', 'd', 'e', 'l', 't', 'a', 'e', 'x', '2', '5', 'A', 'T', 'P', 'b', 'i', 'n', 'd', 'i
'1', 'd', 'e', 'l', 't', 'a', 'e', 'x', '2', '5'], ['2', '6', 'A', 'T', 'P', 'b', 'i', 'n', 'd', 'i', 'n', 'g'], ['1', 'L', 'T', 'C', '4
'o', 'r', 't', 'e', 'r', 'l', 'e', 'u', 'k', 'o', 't', 'r', 'i', 'e', 'n', 'e'], ['N', 'U', 'P', '2', '1', '4'], ['p', 'r', 'o', 'r', 'e], [
'A', 'N'], ['p', 'u', 't', 'a', 't', 'i', 'v', 'e'], ['o', 'n', 'c', 'o', 'g', 'e', 'n', 'e', 'n', 'u', 'c', 'l', 'e', 'o', 'p', 'o', 'r
'D', 'a'], ['F', 'U', 'S'], ['R', 'N', 'A', 'b', 'i', 'n', 'd', 'i', 'n', 'g'], ['F', 'U', 'S', '7', '5'], ['D', 'N', 'A', 'p', 'a', 'i
'o', 't', 'e', 'i', 'n', 'F', 'U', 'S', 'E', 'R', 'G'], ['c', 'h', 'i', 'm', 'e', 'r', 'i', 'c'], ['p', 'r', 'o', 't', 'e', 'i', 'n', 'f
'p', 'r', 'o', 't', 'e', 'i', 'n', 'f', 'u', 's', 'e', 'd'], ['s', 'a', 'r', 'c', 'o', 'm', 'a', 'f', 'u', 's', 'i', 'o', 'n'], ['m', 'y
'p', 'o', 's', 'a', 'r', 'c', 'o', 'm', 'a', 'e', 't', 'e', 'r', 'o', 'g', 'e', 'n', 'e', 'o', 'u', 's'], ['r', 'i', 'b', 'o', 'n,
'o', 't', 'e', 'i', 'n'], ['P', '2', 'o', 'n', 'c', 'o', 'g', 'e', 'n', 'e'], ['F', 'U', 'S', 'o', 'n', 'c', 'o', 'g', 'e', 'n', 'e'], [
's', 'l', 'o', 'c', 'a', 't', 'e', 'd'], ['l', 'i', 'p', 'o', 's', 'a', 'r', 'c', 'o', 'm', 'a'], ['K', 'L', 'F', '4'], ['K', 'r', 'u,
'k', 'e'], ['4', 'K', 'r', 'u', 'p', 'e', 'l', 'l', 'i', 'k', 'e'], ['g', 'u', 't'], ['K', 'r', 'u', 'p', 'e', 'l', 'l', 'i,
'i', 'n', 'e', 'p', 'i', 't', 'h', 'e', 'l', 'i', 'l'], ['E', 'Z', 'F', 'g', 'u', 't'], ['f', 'a', 'c', 't', 'o', 'r', 'g', 'u', 't
'd'], ['K', 'r', 'u', 'e', 'p', 'e', 'l', 'l', 'i', 'k', 'e'], ['H', 'L', 'A', 'D', 'P', 'B', '1'], ['D', 'P'], ['W', '4'], ['c', 'h
'D', 'P'], ['s', 'u', 'b', 'u', 'n', 'i', 't', 'M', 'H', 'C'], ['D', 'P', 'B', '1', 'M', 'H', 'C'], ['H', 'L', 'A', 'D', 'P', 'B', 'e,
'P', 'B', '1'], ['H', 'D', 'A', 'C', '2'], ['2', 'Y', 'Y', '1', 'a', 's', 's', 'o', 'c', 'i', 'a', 't', 'e', 'd'], ['1', 't', 'r', 'a', 'i
'i', 'o', 'n', 'a', 'l'], ['R', 'P', 'D', '3'], ['R', 'O', 'C', 'K', '1'], ['r', 'h', 'o', 'a', 's', 's', 'o', 'c', 'i', 'a', 't', 'e,
'R', 'O', 'C', 'K', '1', 'r', 'e', 'n', 'a', 'l'], ['N', 'Y', 'R', 'E', 'N', '3', '5'], ['D', 'N', 'M', 'T', '3', 'B'], ['3', 'B', 'D,
'D', 'N', 'A'], ['H', 'I', 's', 'a', 'I', 'B', 'D', 'N', 'A'], ['H', 's', 'a', 'I', 'B'], ['F', '1', '3', 'A', '1'], ['C
'i', 'o', 'n'], ['X', 'I', 'I', 'I'], ['c', 'h', 'a', 'i', 'n', 'F', 'S', 'F'], ['s', 'u', 'b', 'u', 'n', 'i', 't', 'T', '3', 'A', 'a', 's',
'2', '1', '1'], ['A', '1'], ['p', 'r', 'o', 'l', 'y', 'p', 'e', 't', 'i', 'd', 'e', 'c', 'o', 'a', 'g', 'u', 'l', 'a', 't', 'i', 'o', 'n
'a', 'c', 't', 'o', 'r'], ['X', 'I', 'I', 'I', 'a', 'f', 'i', 'b', 'r', 'i', 'n'], ['s', 't', 'a', 'b', 'i', 'l', 'i', 'z', 'i', 'n', 'g
't', 'f', 'i', 'b', 'r', 'i', 'n', 'g', 'a', 's', 'e', 'p', 'r', 'o', 't', 'e', 'i', 'n', 'g', 'l', 'u', 't', 'a', 'm', 'a,
'a', 'g', 'l', 'u', 't', 'a', 'm', 'y', 'l', 't', 'r', 'a', 'n', 's', 'f', 'e', 'r', 'a', 's', 'e], ['c', 'h', 'a', 'i', 'n', 't', 'r,
'a', 'm', 'i', 'n', 'a', 's', 'e], ['p', 'l', 'a', 's', 'm', 'a'], ['I', 'T', 'G', 'A', '2'], ['a', 'l', 'p', 'h', 'a', '2', 'C', 'D,

[a], [V', 'L', 'A', '2'], [r', 'e', 'c', 'e', 'p', 't', 'o', 'r', 'c', 'o', 'l', 'l', 'a', 'g', 'e', 'n'], [r', 'e', 'c', 'e', 'p', 't', 'n'], [a', 'l', 'l', 'o', 'a', 'n', 't', 'i', 'g', 'e', 'n'], [s', 'y', 's', 't', 'e', 'm'], [5', 'i', 'n', 't', 'e', 'g', 'n', 'i', 'B', 'r', 'n', 'p', 'l', 'a', 't', 'e', 'l', 'e', 't'], [G', 'P', 'I', 'a', 'p', 'l', 'a', 't', 'e', 'l', 'e', 't'], [I', 'a', 'v', 'e', 'I', 'i', 'n', 't', 'e', 'l', 'e', 'u', 'k', 'i', 'n', '2', '3'], [r', 'e', 'c', 'e', 'p', 't', 'o', 'r', 'I', 'L', '2', '3'], [X', 'P', 'm', 'u', 't', 'a', 'n', 't'], [x', 'e', 'r', 'o', 'd', 'e', 'r', 'm', 'a'], [C', 'x', 'e', 'r', 'o', 'd', 'e', 'r', 'm', 'a'], [C', 'c', 'a', 'u', 'd', 'a', 'l'], [2', 'c', 'a', 'u', 'd', 'a', 'l', 't', 'y', 'p', 'e'], [m', 'i', 'n', 'i', 'C', 'D', 'X', '2'], [E', 'n', 't', 'e', 'r', 'f', 'e', 'r', 'o', 'n', 'i', 'n', 'd', 'u', 'c', 'e', 'd'], [d', 'o', 'u', 'b', 'l', 'e', 's', 't', 'r', 'a', 'n', 'a', 'c', 't', 'i', 'v', 'e', 't', 'e', 'd'], [k', 'i', 'n', 'a', 's', 'e', 'e', 'I', 'e'], [s', 't', 'r', 'a', 'n', 'd', 'e', 'd'], [a', 'c', 't', 'i', 'v', 'e', 't', 'e', 'd'], [k', 'i', 'n', 'a', 's', 'e', 'e', 'I', 'e', 'r', 'f', 'e', 'r', 'o', 'n', 'i', 'n', 'd', 'u', 'c', 'e', 'd'], [e', 'l', 'F', '2', 'a', 'l', 'p', 'h', 'a'], [k', 'i', '8'], [R', 'p', 'n', 'o', 't', 'e', 'i', 'n'], [i', 'n', 't', 'e', 'r', 'f', 'e', 'r', 'o', 'n', 'i', 'n', 'd', 'u', 'c', 'e', 'd', 'b', 'l', 'd', 'e', 'n', 't', 'p', 'r', 'o', 't', 'e', 'i', 'n'], [8', '3', 't', 'y', 'r', 'o', 's', 'i', 'n', 'e', 'p', 'r', 'o', 't', 'e', 'i', 'x', 'p', 'o', 'r', 't', 'i', 'n', 'l', 'c', 'h', 'r', 'o', 'm', 'o', 's', 'o', 'm', 'e', 'e'], [h', 'o', 'm', 'o', 'l', 'o', 'g', 'c', 'h', 'e'], [h', 'o', 'm', 'o', 'l', 'o', 'g', 'e', 'x', 'p', 'o', 'r', 't', 'i', 'n'], [C', 'R', 'M', '1'], [e', 'x', 'p', 'o', 'r', 't', 'i', 'r', 'e', 'd'], [G', 'L', 'I', '1'], [G', 'L', 'I', '1', 'G', 'L', 'I', 'K', 'r', 'u', 'p', 'p', 'e', 'l'], [G', 'L', 'I', '1', 's', 's', 'o', 'c', 'i', 'a', 't', 'e', 'd'], [1', 'g', 'l', 'i', 'o', 'm', 'a', 'a', 's', 'o', 'c', 'i', 'a', 't', 'e', 'd'], [G', 'T', '1'], [h', 'e', 'p', 'c', 'a', 'r', 'c', 'i', 'n', 'l', 'o', 'n', 'g'], [4', '7', 'm', 'e', 't', 'a', 's', 't', 'a', 's', 'i', 'e', 'n', 'o', 'c', 'a', 'r', 'c', 'i', 'n', 'o', 'm', 'a'], [e', 'r', 'i', 'c', 'h', 'e', 'd'], [a', 'b', 'u', 'n', 'd', 'a', 'n', 'e', 'a', 'r', 'r'], [p', 'a', 'r', 'a', 's', 'p', 'e', 'c', 'k', 'l', 'e'], [a', 's', 's', 'e', 'm', 'b', 'l', 'y'], [T', 'F', 'R', 'C'], [r', 'r', 'i', 'n'], [E', 'P', 'H', 'X', '1'], [e', 'p', 'o', 'x', 'i', 'd', 'e'], [1', 'e', 'p', 'o', 'x', 'i', 'd', 'e'], [h', 'y', 'e', 'p', 'o', 'x', 'i', 'd', 'e'], [m', 'i', 'c', 'r', 'o', 's', 'o', 'r', 'o', 'm', 'a'], [a', 'l', 'e', 'p', 'o', 'x', 'i', 'd', 'e'], [m', 'i', 'c', 'l', 'x', 'e', 'n', 'o', 'b', 'i', 'o', 't', 'i', 'c', 'l], [M', 'U', 'T', 'Y', 'H], [a', 'd', 'e', 'n', 'i', 'n', 'e'], [g', 'l', 'y', 'e', 'A', 'G', 's', 'p', 'e', 'c', 'i', 'f', 'i', 'c], [g', 'l', 'y', 'c', 'o', 's', 'y', 'l', 'a', 's', 'e', 'm', 'u', 't', 'Y'], [h', 'u', 't', 'Y', 'l', 'i', 'k', 'e'], [P', 'T', 'P', 'R', 'C], [C', 'C', 'D', '4', '5'], [a', 'n', 't', 'i', 'g', 'e', 'n', 'T', '2', 'o', 'p', 'r', 'o', 't', 'e', 'i', 'n', 'T', '2', '0', '0'], [C', 'o', 'm', 'o', 'n'], [a', 'n', 't', 'i', 'g', 'e', 'n', 'p', 'r', 'o', 't', 'K', 'X', '2', '1'], [N', 'k', 'x', '2', '1', 'N', 'K', '2'], [A', 'h', 'o', 'm', 'e', 'o', 'b', 'o', 'x'], [A', 't', 'h', 'y', 'r', 'r', 'o', 'i', 'd', 's', 'p', 'e', 'c', 'i', 'f', 'i', 'c], [M', 'I', 'R', '1', '2', '5', 'B', '2], [h', 's', 'a', 'm', 'i', 'n', '1', 'K], [B', 'T', 'K', 'B', 'c', 'e', 'l', 'l], [K', 'i', 'n', 'a', 's', 'e', 'B', 'r', 'u', 't', 'o', 'n], [a', 'g', 'a', 'm', 'i', 'n', 'e', 'a', 'i', 'a'], [k', 'i', 'n', 'a', 's', 'e', 'a', 'g', 'a', 'm', 'i', 'n', 'e', 'a', 'g', 'l', 'o', 'b', 'u', 'l', 'i', 'n', 'a', 'n', 'a', 's', 'e', 'd', 'o', 'm', 'i', 'n', 'a', 'n', 't', 'n', 'e', 'g', 'a', 't', 'i', 'v', 'e'], [k', 'i', 'n', 'a', 's', 'e', 'd', 't', 'l], [B', 'r', 'u', 't', 'o', 'n', 's'], [k', 'i', 'n', 'a', 's', 'e', 't', 'r', 'u', 'n', 'c', 'a', 't', 'e', 'd], [B', 'r', 'u', 's', 'e', 't', 'y', 'r', 'o', 's', 'i', 'n', 'e', 'p', 'r', 'o', 't', 'e', 'i', 'n], [N', 'E', 'S], [n', 'e', 's', 't', 'i', 'n], [C', 'e', 'g', 'a', '2', 'G', 'S', 'T', 'O', '2', 'M', 'M', 'A], [r', 'e', 'd', 'u', 'c', 't', 'a', 's', 'e', 'b', 'A', '1', '2', '7', 'L', 't', 'a', 't', 'h', 'i', 'o', 'n', 'e', 's', 't', 'r', 'a', 'n', 's', 'f', 'e', 'r', 'a', 's', 'e'], [o', 'm', 'e', 'g', 'a'], [2', '2', 'h', 'i', 'o', 'n', 'e', 's', 't', 'r', 'a', 'n', 's', 'f', 'e', 'r', 'a', 's', 'e', 'l', 'i', 'k', 'e'], [p', 'r', 'o', 't', 'e', 'i', 'u', 'c', 't', 'a', 's', 'e', 'm', 'o', 'n', 'o', 'm', 'e', 't', 'h', 'y', 'l', 'a', 'r', 's', 'o', 'n', 'i', 'c], [E', 'L', 'N], [e', 't', 'r', 'o', 'p', 'o', 'e', 'l', 'a', 's', 't', 'i', 'n], [I', 'D', 'H', '2], [m', 'i', 't', 'o', 'c', 'h', 'o', 'n', 'd', 'r', 'i', 'I', 'C', 'D', 'H', 'i', 's', 'o', 'c', 'i', 't', 'r', 'a', 't', 'e'], [m', 'i', 't', 'o', 'c', 'h', 'o', 'n', 'd', 'r', 'i', 'c', 'c', 'i', 'n', 'a', 't', 'e], [C', 'D', '2', '4], [C', 'D', '2', '4', 'C', 'D', '2', '4], [I', 'T', 'G', 'A', 'M], [a', 'l', '1', '1], [B', 'C', 'R', '3], [C', 'h', 'a', 'i', 'n', 'a', 'n', 't', 'i', 'g', 'e', 'n], [C', 'D', '1', '1', 'b], [p', '1', '7', '0', 'p', 'h', 'a', 'c', 'o', 'm', 'p', 'l', 'e', 'm', 'e', 'n', 't], [c', 'o', 'm', 'p', 'l', 'e', 'm', 'e', 'n', 't], [M', 'O', '1', 'm', 'a', 'g', 'e], [p', 'o', 'l', 'y', 'p', 'e', 'p', 't', 'i', 'd', 'e', 'n', 'e', 'u', 't', 'r', 'i', 'p', 'h', 'i', 'l', 'l], [a', 'd', 'h', 'a', 'l', 'p', 'h', 'a', 'M], [S', 'T', 'A', 'T', '5', 'A], [5', 'A', 'e', 'p', 'i', 'd', 'i', 'd', 'y', 'm', 'i', 's], [M', 'T', '1', 'a', 'm', 'i', 'n], [c', 'o', 'b], [a', 'l', 'a', 'm', 'i', 'n], [r', 'e', 'd', 'u', 'c', 'i', 'n', 'g], [I', 'L', '7], [N', 's', 'a', 'm', 'i', 'n', '2', '1], [I', 'L', '2', '7], [i', 'n', 't', 'e', 'r', 'l', 'e', 'u', 'k', 'i', 'n', '2', '7], [a', '7], [p', '2', '8], [s', 'u', 'b', 'u', 'n', 'i', 't', 'I', 'L', '2', '7], [a', 'l', 'p', 'h', 'a', 'I', 'L', '2', '7', 'A', 'I', 'e', 'r', 'l', 'e', 'u', 'k', 'i', 'n', '3', '0], [K', 'D', 'M', '1', 'A], [1', 'y', 's', 'i', 'n', 'e', 's', 'p', 'e', 'c', 'i', 'f', 'h', 'y', 'l', 'a', 's', 'e], [1', 'A', 'B', 'R', 'A', 'C], [B', 'H', 'C', '1', '1', '0', 'F', 'A', 'n', 'g], [B', 'R', 'A', 'F', '3', '5', 'H', 'D', 'A', 'C], [s', 'u', 'b', 'u', 'n', 'i', 't', 'a', 'm', 'i', 'n', 'e], [F', 'l', 'a', 'v', 'i', 'n', 'c', 'o', 'n', 'i', 'n', 'g], [a', 'm', 'i', 'n', 'e], [1', 'A', 'l', 'y', 's', 'i', 'n', 'e', 'c', 'l], [C', 'T', 'C', 'F], [C', 'F', '1', '1], [r', 'e', 'p', 'r', 'e', 's', 's', 'o', 'r', '1', '1', 'z', 'i', 'n', 'c], [C', 'C', 'C', 'T', 'C', 'b', 'i', 'n', 'd', 'i', 'n', 'g], [C', 'T', 'C', 'F', 'L], [C', 'X', 'C', 'R', '3], [3', 'G], [9', 'I', 'e',

'p', 'h', 'a', 'c', 'h', 'e', 'm', 'o', 'k', 'i', 'n', 'e', 'l', ['p', 'r', 'o', 't', 'e', 'i', 'n', 'm', 'e', 'l', 'a', 'n', 'o', 'm', 'a']
'e', 'l', 'a', 'n', 'o', 'm', 'a', 'l', ['s', 't', 'i', 'm', 'u', 'l', 'a', 't', 'o', 'r', 'y'], ['a', 'l', 'p', 'h', 'a', 'n', 'e', 'u', 't',
'a', 'c', 't', 'i', 'v', 'a', 't', 'i', 'n', 'g'], ['I', 'G', 'H', 'M'], ['c', 'o', 'n', 's', 't', 'a', 'n', 't'], ['I', 'g', 'M', 'i',
'o', 'b', 'u', 'l', 'l', 'i', 'n'], ['m', 'u'], ['N', 'A', 'T', '1'], ['1', 'N', 'a', 'c', 'e', 't', 'y', 'l', 't', 'r', 'a', 'n', 's', 'f',
'r', 'y', 'l', 'a', 'm', 'i', 'd', 'e'], ['1', 'm', 'o', 'n', 'o', 'm', 'o', 'r', 'c', 'p', 'h', 'i', 'c'], ['F', 'T', '4'], ['3', 'f', 'r',
't', 'y', 'r', 'o', 's', 'i', 'n', 'e', 'p', 'r', 'o', 't', 'e', 'i', 'n'], ['M', 'M', 'P', '8'], ['c', 'o', 'l', 'l', 'a', 'g', 'e', 'n',
'c', 'o', 'l', 'l', 'a', 'g', 'e', 'n', 'a', 's', 'e', 'p', 'M', 'N', 'L'], ['2', 'm', 'a', 't', 'r', 'i', 'x'], ['m', 'e', 't', 'a', 'e', 'i', 'n', 'a', 's', 'e', '8'], ['T', 'A', 'C', 'R', '1'], ['s', 'u', 'b', 's', 't', 'a', 'n', 'c', 'e', 'P'], ['r', 'e', 'c', 'e',
'r', 'e', 'c', 'e', 'p', 't', 'o', 'r', 'N', 'K', '1', 'R', 't', 'a', 'c', 'h', 'y', 'k', 'i', 'n', 'i', 'n'], ['s', 'u', 'b', 's', 't', 'e',
'u', 'r', 'o', 'k', 'i', 'n', 'i', 'n', '1'], ['M', 'I', 'R', '1', '4', '3'], ['h', 's', 'a', 'm', 'i', 'r', '1', '4', '3'], ['X', 'P',
'e', 'r', 'o', 'd', 'e', 'r', 'i', 'a'], ['A', 'C', 'o', 'm', 'p', 'l', 'e', 'm', 'e', 'n', 't', 'i', 'n', 'g'], ['p', 'r', 'o', 't', 'e', 'e',
'e', 'r', 'm', 'a'], ['H', 'S', 'D', '1', '1', 'B', '2'], ['c', 'o', 'r', 't', 'i', 'c', 'o', 's', 't', 'e', 'r', 'o', 'i', 'd'], ['1', 'h',
'h', 'y', 'd', 'r', 'o', 'g', 'e', 'n', 'a', 's', 'e'], ['2', 'H', 'S', 'D', '1', '1'], ['I', 'I', '1', '1', 'D', 'H', '2', '1', '1', 'H
'b', 'e', 't', 'a', 'H', 'S', 'D'], ['I', 'I', '1', '1', 'b', 'e', 't', 'a', 'H', 'S', 'D', '2', '1', '1', 'b', 'e', 't', 'a', 'h', 'y',
'e', 'r', 'o', 'i', 'd'], ['2', '1', '1', 'b', 'e', 't', 'a', 'h', 'y', 'd', 'r', 'o', 'x', 'y', 's', 't', 'e', 'r', 'o', 'i', 'd'], ['I
'p', 'e', 'n', 'd', 'e', 'n', 't'], ['1', '1', 'b', 'e', 't', 'a', 'h', 'y', 'd', 'r', 'o', 'x', 'y', 's', 't', 'e', 'r', 'o', 'i', 'd]
'o', 'g', 'e', 'n', 'a', 's', 'e', 's', 'h', 'o', 'r', 't'], ['d', 'e', 'h', 'y', 's', 'd', 'r', 'o', 'g', 'e', 'n', 'a', 's', 'e', 'r', 'e',
'9', 'C'], ['T', 'E', 'R', 'C'], ['C', 'a', 'j', 'a', 'l'], ['b', 'o', 'd', 'y', 's', 'p', 'e', 'c', 'i', 'f', 'i', 'c'], ['Z', 'E', 'B',
'i', 'n', 'd', 'i', 'n', 'g'], ['1', 'S', 'm', 'a', 'd', 'i', 'n', 't', 'e', 'r', 'a', 'c', 't', 'i', 'n', 'g'], ['1', 'z', 'i', 'n', 'c
'T'], ['n', 'u', 'c', 'l', 'e', 'o', 't', 'i', 'd', 'y', 'l', 'e', 'x', 'o', 't', 'i', 'n', 'a', 'n', 's', 'f', 'e', 'r', 'a', 's', 'e', 'n',
'd', 'e', 't', 'r', 'i', 'p', 'h', 'o', 's', 'p', 'h', 'a', 't', 'e'], ['d', 'e', 'o', 'x', 'y', 'n', 'u', 'c', 'l', 'e', 'o', 't', 'i',
'r', 'a', 'n', 's', 'f', 'e', 'r', 'a', 's', 'e', 't', 'e', 'r', 'm', 'i', 'n', 'a', 'l'], ['a', 'd', 'd', 'i', 't', 'i', 'o', 'd'], ['e
'e', 'r', 'm', 'i', 'n', 'a', 'l'], ['d', 'e', 'o', 'x', 'y', 'n', 'u', 'c', 'l', 'e', 'o', 't', 'i', 'd', 'y', 'l', 't', 'r', 'a', 'n',
't', 'e', 'r', 'm', 'i', 'n', 'a', 'l'], ['d', 'e', 'o', 'x', 'y', 'r', 'i', 'b', 'o', 'n', 'u', 'c', 'l', 'e', 'o', 't', 'i', 'd', 'y,
'e', 'r', 'a', 's', 'e', 't', 'e', 'r', 'm', 'i', 'n', 'a', 'l'], ['L', 'I', 'F'], ['f', 'a', 'c', 't', 'o', 'r', 'D'], ['f', 'a', 'c',
'i', 'n', 'e', 'r', 'g', 'i', 'c'], ['f', 'a', 'c', 't', 'o', 'r', 'd', 'i', 'f', 'f', 'e', 'r', 'e', 'n', 't', 'i', 'a', 't', 'i', 'o',
'i', 't', 'y', 'd', 'i', 'f', 'f', 'e', 'r', 'e', 'n', 't', 'i', 'a', 't', 'i', 'o', 'n', 'i', 'n', 'd', 'u', 'c', 'i', 'n', 'g'], ['f',
'f', 'f', 'e', 'r', 'e', 'n', 't', 'i', 'a', 't', 'i', 'o', 'n', 's', 't', 'i', 'a', 't', 'i', 'n', 'g'], ['f', 'a', 'c',
't', 'o', 'c', 'y', 't', 'e', 's', 't', 'i', 'm', 'u', 'l', 'a', 't', 'i', 'n', 'g'], ['I', 'I', 'I', 'h', 'u', 'm', 'a', 'n'], ['i', 'n
'k', 'i', 'n'], ['D', 'A'], ['c', 'e', 'l', 'l', 's', 'm', 'e', 'l', 'a', 'n', 'o', 'm', 'a', 'd', 'e', 'r', 'i', 'v', 'e', 'd'], ['L,
'2'], ['2', 'N', 'o', 't', 'c', 'h'], ['2', 'n', 'o', 't', 'c', 'h'], ['S', 'P', 'I', '1'], ['P', 'U', '1', '3', '1'], ['p', 'r', 'o', 't
'a', 't', 'o', 'p', 'i', 'e', 't', 'i', 'c'], ['P', 'U', '1', 's', 'p', 'l', 'e', 'e', 'n'], ['f', 'o', 'c', 'u', 's'], ['S', 'F',
'A', 'R', 'N', 'T'], ['a', 'n', 'y', 'l'], ['t', 'r', 'a', 'n', 's', 'l', 'o', 'c', 'a', 't', 'o', 'r', 'c', 'l', 'a', 's', 's'], ['2',
't', 'r', 'a', 'n', 's', 'l', 'o', 'c', 'a', 't', 'o', 'r', 'h', 'y', 'p', 'o', 'x', 'i', 'a', 'i', 'n', 'd', 'u', 'c', 'i', 'b', 'l',
'X', 'R', 'C', 'C', '4', 'X', 'n', 'a', 'y'], ['G', 'A', 'S', '6'], ['a', 'r', 'n', 'e', 's', 't', 's', 'p', 'e', 'c', 'i', 'f', 'i',
'i', 'g', 'a', 'n', 'd', 'A', 'X', 'L'], ['L', 'M', 'O', '1'], ['r', 'h', 'o', 'm', 'b', 'o', 't', 'i', 'n', '1', 'L', 'I', 'M'], ['r',
'n'], ['L', 'I', 'M'], ['1', 'L', 'M', 'O', '1', 'T', 'C', 'e', 'l', 'l'], ['1', 'c', 'y', 's', 't', 'e', 'i', 'n', 'e', 'r', 'i', 'c',
'D', 'A', 'C', '4'], ['4', 'h', 'i', 's', 't', 'o', 'n', 'e'], ['M', 'U', 'C', '4'], ['m', 'u', 'c', 'i', 'n', '4', 'a', 's', 'c', 'i,
'o', 'g', 'l', 'y', 'c', 'o', 'p', 'r', 'o', 't', 'e', 'i', 'n', 'm', 'u', 'c', 'i', 'n'], ['t', 'r', 'a', 'c', 'h', 'e', 'o', 'b', 'r',
'p', 'a', 'n', 'c', 'r', 'e', 'a', 't', 'i', 'c'], ['m', 'u', 'c', 'i', 'n', 't', 'e', 's', 't', 'i', 's'], ['m', 'u', 'c', 'i', 'n', 't
'b', 'r', 'o', 'n', 'c', 'h', 'i', 'a', 'l'], ['m', 'u', 'c', 'i', 'n'], ['M', 'L', 'L', 'T', '1'], ['E', 'N', 'L', 'C', 'T', 'C', '5',
'L', 'M', 'L', 'L'], ['f', 'u', 's', 'i', 'o', 'n', 'M', 'L', 'L', 'E', 'N', 'L'], ['p', 'r', 'o', 't', 'e', 'i', 'n', 'M', 'L', 'L', 'T
's', 'i', 'o', 'n', 'Y', 'E', 'A', 'T', 'S'], ['E', 'N', 'O', '2'], ['g', 'a', 'm', 'm', 'a', 'e', 'n', 'o', 'l', 'a', 's', 'e', '2', 'p
'D', 'g', 'l', 'y', 'c', 'e', 'r', 'a', 't', 'e'], ['h', 'y', 'd', 'r', 'o', 'l', 'y', 'a', 's', 'e', '2', 'p', 'h', 'o', 's', 'p', 'h,
'r', 'a', 't', 'e'], ['h', 'y', 'd', 'r', 'o', 'l', 'y', 'a', 's', 'e', 'n', 'o', 'l', 'a', 's', 'e'], ['n', 'e', 'u', 'r', 'o', 'n
'i', 'd', 'y', 'm', 'i', 's'], ['2', '7', '9', 'n', 'e', 'u', 'r', 'a', 'l'], ['e', 'n', 'o', 'l', 'a', 's', 'e', 'e', 'n', 'e', 'u', 'r', 'o
's', 'e', 'n', 'e', 'u', 'r', 'o', 'n', 's', 'p', 'e', 'c', 'i', 'f', 'i', 'c'], ['e', 'n', 'o', 'l', 'a', 's', 'e', 'n', 'e', 'u', 'r,
'o', 'l', 'a', 's', 'e', 'n', 'e', 'u', 'r', 'o', 'n', 'e', 's', 'p', 'e', 'c', 'i', 'f', 'i', 'c'], ['e', 'n', 'o', 'l', 'a', 's', 'e', 'n', 'e', 'u', 'r,
'I', 'A', 'P'], ['B', 'I', 'A', 'P', '2', 'N', 'F', 'R', '2', 'T', 'R', 'A', 'F'], ['s', 'i', 'g', 'n', 'a', 'l', 'i', 'n', 'g'], ['
'R', 'I', 'N', 'G'], ['4', '8', 'R', 'I', 'N', 'G', 't', 'y', 'p', 'e'], ['B', 'I', 'R', 'C', '2', 'T', 'N', 'F', 'R', '2', 'T', 'R', 'A
'l', 'i', 'n', 'g'], ['2', 'a', 'p', 'o', 'p', 't', 'o', 's', 'i', 's'], ['1', 'c', 'e', 'l', 'l', 'u', 'l', 'a', 'r'], ['1', 'i', 'n',
'A', 'R', 'B', '1'], ['b', 'e', 't', 'a', 'a', 'r', 'n', 'e', 's', 't', 'i', 'n', '1', 'a', 'r', 'n', 'e', 's', 't', 'i', 'n'], ['
'u', 'a', 'l'], ['a', 'r', 'n', 'e', 's', 't', 'i', 'n', '2'], ['N', 'P', 'C', '1'], ['N', 'P', 'C'], ['i', 'n', 't', 'r', 'a', 'c', 'e
'c', 'h', 'o', 'l', 'e', 's', 't', 'e', 'r', 'o', 'l'], ['1', 'N', 'i', 'e', 'm', 'a', 'n', 'n', 'P', 'i', 'c', 'k'], ['C', '1'], ['p,
'r', 'u', 'n', 'c', 'a', 't', 'e', 'd'], ['N', 'i', 'e', 'm', 'a', 'n', 'n', 'P', 'i', 'c', 'k'], ['M', 'I', 'R', '1', '9', '6', 'A', '2
'1', '9', '6', '2', 'h', 's', 'a', 'm', 'i', 'r', '1', '9', '6', 'a', '2', 'm', 'i', 'c', 'r', 'o', 'R', 'N', 'A'], ['1', '9', '6', '2]
'y', 'm', 'p', 'h', 'o', 'c', 'y', 't', 'e', 's', 'p', 'e', 'c', 'i', 'f', 'i', 'c'], ['f', 'a', 'c', 't', 'o', 'r', 'm', 'u', 'l', 't,
'e', 'l', 'o', 'm', 'a'], ['M', 'I', 'R', '1', '2', '4', '1'], ['h', 's', 'a', 'm', 'i', 'r', '1', '2', '4', '1', 'h', 's', 'a', 'm, i
'm', 'i', 'c', 'r', 'o', 'R', 'N', 'A'], ['1', '2', '4', 'a', 'l'], ['L', 'R', 'P', '6'], ['l', 'o', 'w', 'd', 'e', 'n', 's', 'i', 't',
'o', 't', 'e', 'i', 'n'], ['6', 'L', 'R', 'P', '6'], ['S', 'A', 'L', 'L', '4'], ['s', 'a', 'l', 'l', 'i', 'k', 'e'], ['4', 'z', 'i', 'n,
'n', 'c'], ['M', 'E', 'D', '1', '2'], ['1', '2', 'C', 'A', 'G'], ['4', 'S', 'O', 'P', 'A', 'c', 'o', 'n', 't', 'a', 'i', 'n', 'i', 'n,
'n', 'a', 'c', 't', 'i', 'v', 'a', 't', 'o', 'r', 'n', 'e', 'c', 'r', 'u', 'i', 't', 'e', 'd'], ['c', 'o', 'f', 'a', 'c', 't', 'o', 'r],
'p', 'o', 'n', 'e', 'i', 'v', 'h', 'u', 'm', 'a', 'n'], ['o', 'p', 'p', 'o', 's', 'i', 't', 'e'], ['p', 'a', 'i', 'r', 'e', 'd', 'm', 'e
'1', '2'], ['h', 'o', 'm', 'o', 'l', 'o', 'g', 'p', 'u', 't', 'a', 't', 'i', 'v', 'e'], ['1', '2', 't', 'h', 'y', 'r', 'o', 'i', 'd],
'r', 'a', 's', 's', 'o', 'c', 'i', 'a', 't', 'e', 'd'], ['2', '3', '0'], ['c', 'o', 'm', 'p', 'o', 'n', 'e', 'n', 't', 't', 'h', 'y', 'r
'u', 'n', 't', 't', 'r', 'i', 'n', 'u', 'c', 'l', 'e', 'o', 't', 'i', 'd', 'e'], ['T', 'H', 'R', 'A', 's', 's', 'o', 'c', 'i', 'a',
'n', 'u', 'c', 'l', 'e', 'o', 't', 'i', 'd', 'e'], ['M', 'I', 'R', '2', '2', '2'], ['h', 's', 'a', 'm', 'i', 'r', '2', '2', '2'], ['D,
'd', 'e', 'p', 'e', 'n', 'd', 'e', 'n', 't'], ['D', 'X', '3', 'X', 'D', 'E', 'A', 'D'], ['A', 's', 'p', 'G', 'l', 'u', 'a', 'l', '1', 'a
'n', 'k', 'e', 'd', 'D', 'E', 'A', 'D'], ['X', 'c', 'h', 'r', 'o', 'm', 'o', 's', 'o', 'm', 'a', 'l', 'D', 'E', 'A', 'D'], ['i', 's', 'o
'A', 'D', 'H'], ['A', 's', 'p', 'G', 'l', 'u', 'a', 'l', 'a', 's', 'p', 'H', 'i', 's'], ['3', 'D', 'E', 'A', 'D', 'H'], ['b', 'o,
'a', 's', 'e', 'l', 'i', 'k', 'e'], ['M', 'Y', 'C', 'B', 'P', '2'], ['M', 'Y', 'C', 'B', 'P', '2', 'H', 'i', 'g', 'h', 'w', 'i', 'r', 'e
'a', 's', 'e', 'P', 'A', 'M', 'H', 'i', 'g', 'h', 'w', 'i', 'r', 'e', 'R', 'P', 'M', '1'], ['M', 'Y', 'C', 'B', 'P', '2', 'm', 'y', 'c,
'2', 'p', 'a', 'm', 'h', 'i', 'g', 'h', 'w', 'i', 'r', 'e', 'r', 'p', 'm', '1'], ['M', 'y', 'c'], ['M', 'I', 'R', '1', '5', '0'], ['h,
'0'], ['M', 'E', 'C', 'O', 'M'], ['M', 'E', 'C', 'O', 'M', 'D', 'S', '1'], ['E', 'V', 'I', '1'], ['E', 'V', 'I', '1', 'A', 'M', 'L
'r', 'o', 't', 'e', 'i', 'n', 'M', 'D', 'S', '1', 'P', 'R'], ['3', 'e', 'c', 'o', 't', 'r', 'o', 'p', 'i', 'c'], [
'm', 'y', 'e', 'l', 'o', 'd', 'y', 's', 'p', 'l', 'a', 's', 'i', 'a'], ['s', 'y', 'n', 'd', 'r', 'o', 'm', 'e', 'a', 's', 's', 'o', 'c,
'o', 'n', 'c', 'o', 'g', 'e', 'n', 'e'], ['E', 'V', 'I', '1', 'z', 'i', 'n', 'c'], ['E', 'v', 'i', '1'], ['H', 'S', 'P', 'A', '1', 'B],
'1', 'H', 'S', 'P', '7', '0', '2', 'H', 'S', 'P', '7', '0', '1', 'H', 'S', 'P', '7', '0', '2', 'H', 'e', 'a', 't'], ['1', 'H', 'e', 'a,
'1', 'B', 'h', 'e', 'a', 't'], ['N', 'C', 'O', 'A', '2'], ['c', 'e', 'l', 'a', 's', 's'], ['7', '5', 'g', 'l', 'u', 'c', 'o', 'c', 'o,
'p', 'r', 'o', 't', 'e', 'i', 'n', '1', 'p', '1', '6', '0'], ['s', 't', 'e', 'r', 'o', 'i', 'd'], ['2', 't', 'r', 'a', 'n', 's', 'c', 'r
'a', 'l'], ['i', 'n', 't', 'e', 'r', 'm', 'e', 'd', 'i', 'a', 'n', 'y'], ['C', 'A', 'P', 'N', '1'], ['c', 'a', 'l', 'p', 'a', 'i', 'n,
't', 'C', 'A', 'N', 'P'], ['1', 'c', 'a', 'l', 'c', 'i', 'u', 'm', 'a', 'c', 't', 'i', 'v', 'a', 't', 'e', 'd'], ['1', 'c', 'a', 'l', 'p
's', 'u', 'b', 'u', 'n', 'i', 't', 'c', 'a', 'l', 'p', 'a', 'i', 'n'], ['m', 'u', 't', 'y', 'p', 'e', 'c', 'a', 'l', 'p', 'a', 'i', 'n
'a', 'i', 'n', '1'], ['s', 'u', 'b', 'u', 'n', 'i', 't', 'c', 'e', 'l', 'l'], ['3', '0'], ['p', 'r', 'o', 't', 'e', 'i', 'n', 'c', 'e,
'r', 'o', 'm', 'o', 'l', 'a', 'r', 'c', 'a', 'l', 'p', 'a', 'i', 'n'], ['G', 'A', 'D', '4', '5'], ['a', 'l', 'p', 'h', 'a', 'D', 'I', '1', 'D', 'N
'i', 'n', 'D', 'N', 'A'], ['t', 'r', 'a', 'n', 's', 'c', 'r', 'i', 'p', 't', '1', 'g', 'r', 'o', 'w', 't', 'h'], ['D', 'N', 'A', 'd', 'a
'd', 'u', 'c', 'i', 'b', 'l', 'e'], ['4', '5'], ['L', 'U', 'N', 'A', 'R', '1'], ['1', 'e', 'u', 'k', 'e', 'm', 'i', 'a', 'i', 'n', 'd,
'1', '6', '1'], ['h', 's', 'a', 'm', 'i', 'r', '1', '6', '1'], ['D', 'H', 'F', 'R'], ['d', 'i', 'h', 'y', 'd', 'r', 'o', 'f', 'o', 'l', 'l,
'r', 'e', 'p', 'l', 'i', 'c', 'a', 't', 'i', 'o', 'n'], ['l', 'i', 'c', 'e', 'n', 's', 'i', 'n', 'g'], ['M', 'C', 'M', '2', 'c', 'e,
'i', 'o', 'n'], ['c', 'y', 'c', 'l', 'e', 'l', 'i', 'k', 'e'], ['1', 'c', 'y', 'c', 'l', 'i', 'n', 'g'], ['1', 'm', 'i,
'o', 's', 'o', 'm', 'e'], ['d', 'e', 'f', 'i', 'c', 'i', 'e', 'n', 't'], ['m', 'i', 't', 'o', 't', 'i', 'n'], ['m', 'i', 'n', 'i', 'c,

[illegible]

g', 'r', 'u', 'g'], ['C', 'D', '3', '3'], ['C', 'D', '3', '3', 'C', 'D', '3', '3'], ['g', 'p', '6', '7'], ['t', 'r', 'a', 'n', 's', 's', '7', 's', 'i', 'a', 'l', 'i', 'c'], ['a', 'c', 'i', 'd', 'b', 'i', 'n', 'd', 'i', 'n', 'g'], ['D', 'C', 'X'], ['d', 'o', 'u', 'b', 'l', 'e', 'd', 'o', 'u', 'b', 'l', 'e', 'c', 'o', 'r', 't', 'e', 'x', 'd', 'o', 'u', 'b', 'l', 'e', 'i', 'n', 'l', 'i', 's', 'X', 'l', 'i', 's', 's', 'l', 'i', 'n', 'X'], ['C', 'S', 'P', 'G', '4'], ['c', 'h', 'o', 'n', 'd', 'u', 'r', 'i', 't', 'i', 'n'], ['4', 'c', 'h', 'o', 'n', 'd', 'e', 'l', 'a', 'n', 'o', 'm', 'a', 'a', 's', 's', 'o', 'c', 'i', 'a', 't', 'e', 'd'], ['N', 'G', '2', 'm', 'e', 'l', 'a', 'n', 'o', 'm', 'g', 'l', 'y', 'c', 'a', 'n', 'm', 'e', 'l', 'a', 'n', 'o', 'm', 'a', 'a', 's', 's', 'o', 'c', 'i', 'a', 't', 'e', 'd'], ['I', 'H', 'H'], ['p', 'r', 'o', 't', 'e', 'i', 'n', 'I', 'n', 'd', 'i', 'a', 'n'], ['G', 'S', 'T', 'O', '1'], ['o', 'm', 'e', 'g', 'a', '1', 'M', 'M', 'a', 's', 'e', 'S'], ['P', 'h', 'e', 'n', 'a', 'c', 'y', 'l'], ['r', 'e', 'd', 'u', 'c', 't', 'a', 's', 'e', 'e', 'p', 'i', 'd', 'i', 'd', 'g', 'l', 'u', 't', 'a', 't', 'h', 'i', 'o', 'n', 'e'], ['1', 'l', 'g', 'l', 'u', 't', 'a', 't', 'h', 'i', 'o', 'n', 'e', 'S', 't', 'r', 's', 'e', 'l', 'l', 'k', 'e', 'g', 'l', 'u', 't', 'a', 't', 'h', 'i', 'o', 'n', 'e', 'd', 'e', 'p', 'e', 'n', 'd', 'e', 'n', 't', 't'], ['V', 'o', 't', 'a'], ['C', 'h', 'a', 'i', 'n', 'C', 'D', '1', '7', '9'], ['A', 'p', 'r', 'i', 'l', 'B'], ['1', 'v'], ['p', 'r', 'i', 'e'], ['S', 'O', 'o', 'l'], ['O', 'a', 'c', 'y', 'l', 't', 'r', 'a', 'n', 's', 'f', 'e', 'r', 'a', 's', 'e'], ['1', 'a', 'c', 'y', 'l', 'c', 'o', 'e', 'n', 'y', 'l', 't', 'r', 'a', 'n', 's', 'f', 'e', 'r', 'a', 's', 'e'], ['1', 's', 't', 'e', 'r', 'o', 'l'], ['a', 'c', 'y', 'l', 'c', 'o', 'e', 'N', 'M', 'T'], ['h', 'i', 'c', 'o', 'i', 'n', 'a', 'm', 'i', 'd', 'e'], ['P', 'O', 'L', 'D', '1'], ['d', 'e', 'l', 't', 'a'], ['s', 'D', 'C', '2'], ['h', 'o', 'm', 'l', 'o', 'g', 'D', 'N', 'A'], ['p', 'l', '2', '5', 'p', 'l', 'y', 'm', 'e', 'r', 'a', 's', 'e', 'e', 'd'], ['1', '2', '5', 'k', 'D', 'a', 'p', 'o', 'l', 'y', 'm', 'e', 'r', 'a', 's', 'e'], ['T', 'H', 'A', 'P', '1'], ['T', 'H', 'A', '3', '1', 'A', '0', '1', 'R', 'i', 'k', 'T', 'H', 'A', 'P'], ['1', 'T', 'H', 'A', 'P'], ['p', 'r', 'o', 'a', 'p', 'o', 'p', 't', 'o', 't', 'l'], ['s', 'p', 'l', 'i', 'c', 'i', 'n', 'g'], ['U', '2', 'A', 'F'], ['3', 'A'], ['s', 'u', 'b', 'u', 'n', 'i', 't', 'U', '2'], ['a', 'y'], ['1', 'U', '2'], ['3', '5', 'K', 'D'], ['s', 'n', 'R', 'N', 'P'], ['a', 'u', 'x', 'i', 'l', 'i', 'a', 'r', 'y'], ['R', 'N', 'U', 'n', 's', 'p', 'l', 'i', 'c', 'i', 'n', 'g'], ['3', '5', 'k', 'D', 'a'], ['F', 'O', 'X', 'O', '4'], ['O', '4', 'f', 'o', 'n', 'k'], ['h', 'l', 'm', 'y', 'e', 'l', 'o', 'i', 'd', 'l', 'y', 'm', 'p', 'h', 'o', 'i', 'd'], ['Z', 'N', 'F', '3', '8', '4'], ['3', '8', '4', 'C', 'A', 'n', 't', 'e', 'r', 'a', 'c', 't', 'i', 'n', 'g'], ['p', 'r', 'o', 't', 'e', 'i', 'n', 'e', 'x', 'p', 'a', 'n', 'd', 'e', 'd'], ['C', 'A', 'n', 'i', 'n', 'u', 'c', 'l', 'e', 'o', 't', 'i', 'd', 'e'], ['M', 'B', 'D', '2'], ['m', 'e', 't', 'h', 'y', 'l', 'C', 'p', 'G', 'b', 'i', 'd', 'e', 'm', 'e', 't', 'h', 'y', 'l', 'a', 's', 'e'], ['M', 'I', 'R', '1', '4', '4'], ['h', 's', 'a', 'm', 'i', 'r', '1', '4', '4'], ['i', 'g', 'e', 'n', 'C', 'D', '2', '7'], ['1', 'i', 'g', 'a', 'n', 'd', 'K', 'i', '2', '4'], ['a', 'n', 't', 'i', 'g', 'e', 'n', 's', 'u', 'D', '7', '0', 't', 'u', 'm', 'o', 'r'], ['8', 'A', 't', 'u', 'm', 'o', 'r'], ['H', 'O', 'X', 'B', '7'], ['H', 'o', 'x', 'B', '7', 'h', 'o', 'm', 'e', 'o'], ['B', '7', 'h', 'o', 'm', 'e', 'o'], ['c', '1'], ['p', 'r', 'o', 't', 'e', 'i', 'n', 'h', 'o', 'm', 'e', 'o', 'b', 'h', 'o', 'm', 'e', 'o', 'b', 'o', 'x'], ['H', 'S', 'P', 'A', '1', 'L'], ['1', 'i', 'k', 'e', 'e', 'p', 's', 's'], ['1', '0', 'k', 'D', 'a'], ['1', 'l', 'i', 'k', 'e', 'h', 'e', 'a', 't'], ['1', 'H', 'o', 'm', 'h', 'e', 'a', 't'], ['1', 'L', 'h', 't', 'e', 'i', 'n', 'l', 'i', 'k', 'e'], ['1', 'l', 'i', 'k', 'e'], ['M', 'E', 'I', 'S', '1'], ['M', 'e', 'i', 's', '1', 'M', 'e', 'i', 'o', 'g', 'W', 'U', 'G', 'S', 'C'], ['H', 'N', 'H', '0', '4', '4', 'B', '0', '4', '1', 'l', 'e', 'u', 'k', 'e', 'm', 'o', 'g', 'M', '2'], ['P', 'R'], ['2', 'G', 'A', 'T', 'A', '3'], ['G', '3', 'B', 'M', 'T', 'E', 'b', 'i', 'n', 'd', 'i', 'n', 'g'], ['p', 'r', 'o', '2', 'P', 'R'], ['Z', 'N', 'F'], ['d', 'o', 'm', 'a', 'i', 'n', 'P', 'R'], ['8', 'r', 'e', 't', 'i', 'n', 'o', 'b', 'l', 'a', 's', 't', 't', 'e', 'i', 'n', 'b', 'i', 'n', 'd', 'i', 'n', 'g'], ['p', 'r', 'o', 't', 'e', 'i', 'n', 'e', 'i', 'n', 'r', 'e', 't', 'i', 'n', 'o', 'b', 'l', 'a', 'r', 'o', 't', 'e', 'i', 'n', 't', 'e', 'r', 'a', 'c', 't', 'i', 'n', 'g'], ['R', 'I', 'Z', 'z', 'i', 'n', 'c', 'f', 'i', 'n', '9', 'X'], ['p', 'r', 'o', 'b', 'a', 'b', 'l', 'e'], ['F', 'A', 'F', 'X', 'D', 'r', 'o', 's', 'o', 'p', 'h', 'i', 'l', 'a'], ['f', 'a', 's', 's'], ['X', 'l', 'i', 'n', 'k', 'e', 'd', 'd', 'e', 'u', 'b', 'i', 'q', 'u', 'i', 't', 'i', 'n', 'a', 't', 'i', 'n', 'g'], ['F', 'A', 'F', 'm', 'a', 'l', 's', 'f', 'a', 't'], ['X', 'l', 'i', 'n', 'k', 'e', 'd', 'h', 'F', 'A', 'M', 'u', 'b', 'i', 'q', 'u', 'i', 't', 'i', 's', 'l', 'i', 'i', 'k', 'e'], ['t', 'h', 'i', 'e', 's', 't', 'e', 'r', 'a', 's', 'e'], ['F', 'A', 'F', 'X', 'u', 'b', 'i', 'q', 'u', 'i', 'o', 'l', 'e', 's', 't', 'e', 'r', 'a', 's', 'e'], ['F', 'A', 'F', 'X', 'u', 'b', 'i', 'q', 'u', 'i', 't', 'i', 'n', 's', 'p', 'e', 'c', 'o', 'c', 'e', 's', 'i', 'n', 'g'], ['K', 'D', 'M', '6', 'A'], ['6', 'A', 'b', 'A', '3', '8', '6', 'N', '1', '4', '2'], ['u', 'b', 's', 'l', 'y'], ['t', 'r', 'a', 'n', 's', 'c', 'r', 'i', 'b', 'e', 'd'], ['t', 'e', 't', 'r', 'a', 't', 'r', 'i', 'c', 'o', 'p', 'e', 'X'], ['U', 'T', 'X', 'l', 'y', 's', 'i', 'n', 'e'], ['6', 'A', 'u', 'b', 'i', 'q', 'u', 'i', 't', 'o', 'u', 's', 'l', 'y'], ['X', 'l', 'i', 'q', 'u', 'i', 't', 'o', 'u', 's', 'l', 'y', 't', 'r', 'a', 'n', 's', 'c', 'r', 'i', 'b', 'e', 'd'], ['T', 'P', 'R'], ['E', 'P', 'S', 'e', 'r', 'm', 'a', 'l'], ['1', '5', 'A', 'L', 'L', '1'], ['f', 'u', 's', 'e', 'd'], ['A', 'F', '1', 'p'], ['P', 'L', 'K', '4'], ['P', 'K', 'i', 'n'], ['1', '8', 's', 'e', 'n', 'i', 'n', 'e', 't', 'h', 'h', 'r', 'e', 'o', 'n', 'i', 'n', 'e', 'p', 'r', 'i', 'o', 't', 'e', 'i', 'n'], ['R', 'A'], ['i', 'n', 't', 'e', 'r', 'l', 'e', 'u', 'k', 'i', 'n', '3'], ['a', 'l', 'p', 'h', 'a', 'C', 'D', '1', '2', '3'], ['a', 'n', '3'], ['S', 'P', '2'], ['i', 's', 'o', 'f', 'o', 'r', 'm', 'I', 'L', '3'], ['a', 'l', 'p', 'h', 'a', 'I', 'L', '3', 'R'], ['a', 'l', 'p', 'l', 'p', 'h', 'a', 'I', 'L', '3', 'R', 'A', 'i', 'n', 't', 'e', 'r', 'l', 'e', 'u', 'k', 'i', 'n'], ['l', 'o', 'w'], ['a', 'f', 'i', 'F', 'R', 'S', 'F', '1', '3', 'C'], ['1', '3', 'C', 'B'], ['c', 'e', 'l', 'a', 'c', 't', 'i', 'v', 'a', 't', 'i', 'n', 'g'], ['r', 'B', 'A', 'F', 'F'], ['r', 'e', 'c', 'e', 'p', 't', 'o', 'r', 'B', 'L', 'y', 'S'], ['B', 'L', 'K'], ['B', 'l', 'k', 'B'], ['k', 'i', 'n', 'k', 'i', 'n', 'a', 's', 'e', 'b'], ['k', 'i', 'n', 'a', 's', 'e', 'p', '5', '5', 'B', 'l', 'k'], ['C', 'D', '2', '2'], ['C',

e', i', n', [A', F', '6'], [T', B', L', '1', X', R', '1'], [F', b', o', x', l', i', k', e', W', D'], [T', B', 'L', '1', 'r', 'e', 'l', 'a', 't', 'e', 'd'], [c', 'o', 'r', 'e', 'p', 'r', 'e', 's', 's', 'o', 'r', 'H', 'D', 'A', 'C', '3'], [T', B', 'L', '1', 'P', 'a', 'x', '4', 'p', 'a', 'i', 'r', 'e', 'd'], [4', 'p', 'a', 'i', 'r', 'e', 'd'], [B', C', L', '1', '1', 'B'], [l', y', m', u', k', e', m', a', i', a'], [l', '1', 'B', B], [l', '1', 'B', B', c', e', l', l'], [l', '1', 'B', T], [l', '1', 'B', T', c', p', h', o', m', a', l', e', u', k', a', e', m', i', a'], [l', '1', '1', 'B', B', C', L', '1', '1', 'B', B', C', L', '1', o', m', p', o', n', e', n', t', B', C', L', '1', '1', B', T', R', D', C'], [f', u', s', i', o', n', C', O', U', r', a', c', t', i', n', g], [l', '2', 'h', R', i', t', '1', r', a', d', i', a', t', i', o', n', i', n', d', u', c', i', '1'], [M', I', R', '1', '2', '8', '1'], [h', 's', a', m', i', r', '1', '2', '8', '1', h', s', a', m', i', r', '1', '2', '8', R', N', A'], [l', '2', '8', 'a'], [P', C', D', H', '1', '0'], [p', r', o', t', o', c', a', d', h', e', r', i', n', '0'], [l', '6', '0', 'S'], [r', i', b', o', s', o', m', a', l'], [L', '1', '0', W', i', l', m', s'], [t', u', m', o', r', d'], [p', r', o', t', e', i', n', l', a', m', i', n', i', n'], [h', o', m', o', l', o', g', l', a', r', g', e', m', o', r', l], [Q', M], [F', L', T', '3', L', G], [f', m', s', r', e', l', a', t', e', d], [l', i', g', a', n', E', F', '2', D], [l', '2', 'M', A', D', S], [A', C', D], [a', d', r', e', n', o', c', o', r', t', i', c', a', s', i', a'], [h', o', m', o', l', o', g', P', O', T', l], [T', I', N', '2', i', n', t', e', r', a', c', t', i', e', i', n', T', I', N', '2'], [l', a', d', r', e', n', o', c', o', r', t', i', c', a', l], [G', R', I', N', '3', A', N', m', e', t', h', y', l', D', a', s', p', a', r', t', a', t', e', l], [s', u', b', t', y', p', e', l], [l', '3', A', g', l', u', t', a', m', a', t', e', l], [l', '3', A', g', l', u', t', a', m', a', t', e', l], [N', m', e', t', h', y', l', a', t', e', l], [l', '3', A], [B', T', L', A], [a', t', t', e', n', u', a', t', o', r', B], [L', I', M', S', l], [l', s', n', t], [a', n', t', i', g', e', n', l', i', k', e', c', o', n', t', a', i', n', i', n', g], [l', L', I', M', p', e], [l', '1', p', a', r', t', i', c', u', l', a', r', l', a', r', l', y], [i', n', t', e', r', e', s', t', i', n', g], [l', i', s', l], [l', '1', r', e', n', a', l], [N', Y', R', E', N', '4', '8', s', e', n', e', s', c', e', n', t], [l', Z', F', o', s', o', m', a', l], [p', r', o', t', e', i', n', X', l', i', n', k', e', d], [T', R', I', B', '2'], [t', p', i', M', '2'], [p', i', m', '2', p', i', m', '2'], [o', n', c', o', g', e', n', e', p', i', m', '2', h', p', o', g', e', n', e], [P', i', m', '2], [t', h', r', e', o', i', n', e], [G', F', I', '1'], [G', f', i', '1', i', n', d', e', p', e', n', d', e', n', c', e', i', n', g', r', o', w', t', h], [i', n', d', e', p', e', n', d', r', e', s', s', o', r', g', r', o', w', t', h], [l', '6', '3'], [E', B', F', '1'], [C', O', E', '1', C', o', l', f', [E', B', F], [l', e', a', r', l', y], [l', o', l', f', a', c', t', o', r', y], [K', D', M', S', A], [j', i], [A', T], [r', i', c', h], [R', B', P', '2', l', i', k', e], [J', A', R', I', D', l', A', j', u', m', D], [l', A', l', y', s', i', n', e], [l', '5', A', r', e', t', i', n', o', b', l', a', s', t', o', m', a], [S', T', b', i', n', d', i', n', g], [S', T', I', L], [S', C', L', i', n', t', e', r', r', u', p', t', i', n', g', n', S', C', L', T', A', L', '1'], [i', n', t', e', r', r', u', p', t', i', n', g], [l', o', c', u', s', T', A', r', u', p', t', i', n', g], [M', I', R', '9', '9', A], [h', s', a', m', i', r', '9', '9', a], [M', S', I', i], [l', '2', m', u', s', a', s', h', i], [l', '2', m', u', s', a', s', h', i', '2'], [M', I', R', '1', '9', '3', B], [h', '3', b], [C', A', '1', '2'], [c', a', r', b', o', n', i', c], [a', n', h', y', d', r', a', s', e], [l', '2', c', e', l], [d', e', h', y', d', r', a', t', a', s', e], [X', I', I', c', a', r', b', o', n', i', c], [d', e', h', y', e', t', u', m', o', r], [H', O', M', R', C', C', '3', '1', '3'], [T', R', D], [p', o', l', y', p', e', p', t', l], [T', C', R', D], [T', E', S], [t', e', s', t', i', n', t', e', s', t', i', s], [M', N', X', '1'], [m', r', o', n], [p', a', n', c', r', e', a', s], [H', B', '9', h', o', m', e', o', b', o', x], [H', B', '9'], [T', w', i', s', t', n', e', l', a', t', e', d], [l', '3', '9', d', e', m', i', s', s', e', x', p', r', e', s', s', e', t], [l', '2', t', w', i', s', t], [l', '2', t', w', i', s', t', r', e', l', a', t', e', d], [D', e', r', m', o', l], [P', d', e', p', e', n', d', e', n', t], [A', T', F', '5', c', A', M', P', d', e', p', e', n', d', e', n', t], [a', n', s', c', r', i', p', t', i', o', n], [A', T', F', X], [S', O', X', '5], [S', O', X', '5', S', R', Y], e', r', m', i', n', i', n', g], [T', R', G], [r', e', a', r', r', a', n', g', i', n', g], [g', a', m', m', l', u', s', t', e', r', T', C', R', G], [P', P', P', '2', R', '5', A], [l', '5', '6], [i', s', o', f', o', r', m', p', a', a', P', P', '2', A], [B', '5', '6', a', l', p', h', a', P', P', '2', A], [P', R', '6', '1', a', l', p', h', a', l', p', a', P', P', '2', A], [B', '5', '6], [P', R', '6', '1], [R', '5], [i', s', o', f', o', r', m', h', a', p', r', o', t', e', i', n], [a', l', p', h', a', p', r', o', t', e', i', n], [i', s', o', f', o', r', n], [i', s', o', f', o', r', n], [l', i', f', u', s', e', d], [p', r', o', t', e', i', n', Y', E', A', T', S], [l', '3', m', y', e', l', o', i', d], [l', d], [S', P', R', E', D', l], [s', p', r', o', u', t', y', e', l', a', t', e', d], [E', V', H', '1], r', e', s', s', o', r], [R', a', s', M', A', P', K], [S', M', Y', D', '2], [S', M', Y', D', '2', S', E', T], i', s', t', o', n', e], [S', M', Y', D', '2', l', y', s', i', n', e], [l', '3', C', z', i', n', c], [M', I', R', m', i', n', l', '9', '6', b], [A', A', T', F], [A', A', T', F', n', b', i', n', i', n', g], [C', h', e', [H', e', l', i', o', s', i', k', a', r', o', s], [l', '2', z', i', n', c], [H', e', l', i', o', s', z', i', n', c [B', C', A', T', '1], [b', r', a', n', c', h', e', d', c', h', a', i', n', a', m', i', n', o', a', c', i', d', a', a', n', s', f', e', r', a', s', e], [c', y', t', o', s', o', l', i', c', b', r', a', n', c', h', e', d', c', i', d], [t', r', a', n', s', a', m', i', n', a', s', e], [c', y', t', o', s', o', l', i', c', b', r', a', n', c', h', e', d', c', i', d], [t', r', a', n', s', a', m', i', n', a', s', e], [c', y', t', o', s', o', l', i', c', b', r', a', n', c', h', e', d', c', i', d], [l', '1', '8'], [B', C', L', '9], [l', '9', B', c', e', l', l], [p', r', o', t', e', i', n', b', c', l', '9', p', r', o', g', l', e', s', s], [B', L', N', K], [l', i', n', k', e', r], [S', H', '2], [d', o', m', a', i', n', B', c', i', v', a', t', i', o', n', B', c', e', l', l], [p', r', o', t', e', i', n', S', c', i], [l', '6', '5], [K', D], y', t', o', p', l', a', s', m', i', c], [R', A', S', G', R', P', '1], [g', u', a', n', y', l', r', e', l', e', R', A', S], [g', u', a', n', y', l], [n', u', c', l', e', o', t', i', d', e', r', e', l', e', a', s', i', n', s', i', n', g], [c', a', l', c', i', u', m], [D', A', G', r', e', g', u', l', a', t', e', d], [I', I', c], i', a', c', y', l', g', l', y', c', e', r', o', l', r', e', g', u', l', a', t', e', d], [I', I', g', u', a', l', A', n', a', s], [R', a', s', G', R', P], [B', A', A', L', C], [p', r', o', t', e', i', n', B', A', A', l], [b', i', n', d', i', n', g', B', A', A', L', C], [M', E', K', K', l], [b', i', n', d', i', n', g', b', l', [N', A', D', d', e', p', e', n', d', e', n', t], [m', a', l', i', c], [m', i', t', o', c', h', o', n', D', M', E', m', a', l', a', t', e', l], [o', x', a', l', o', a', c', e', t', a', t', e', d', e', c', a', r', b', n', g], [m', i', t', o', c', h', o',

'I', 'G', 'H', 'T', 'l', 'i', 'k', 'e', ['I', 'g', 'H', ['t', 'r', 'a', 'n', 's', 'c', 'r', 'i', 'p', 't', 'i', 'o', 'n', 'E', '2', 'F',
'g'], ['1', 'd', 'e', 'a', 'd', ['r', 'i', 'n', 'g', 'e', 'r', 'l', 'i', 'k', 'e', ['I', 'M', 'P', 'A', '2', ['i', 'n', 'o', 's', 'i',
'o', 'p', 'h', 'o', 's', 'p', 'h', 'a', 't', 'a', 's', 'e', ['2', 'I', 'M', 'P', 'A', 's', 'e', ['2', 'i', 'n',
'n', 'o', 's', 'i', 't', 'o', 'l', ['1', 'i', 'n', 'o', 's', 'i', 't', 'o', 'l', ['m', 'y', 'o', ['2', 'm', 'y', 'o', 'i', 'n', 'o', 'i',
'C', 'A', '2', ['A', 'B', 'C', '1', ['F', 'C', 'R', 'L', '5', ['5', 'f', 'c', ['5', 'f', 'c', 'R', 'l', 'i', 'k', 'e', ['5', 'i', 'r',
'm', 'm', 'u', 'n', 'o', 'g', 'l', 'o', 'b', 'u', 'l', 'i', 'n', ['I', 'R', 'T', 'A', '2', ['C', 'C', 'A', 'R', '1', ['1', 'd', 'e',
'c', 'e', 'r', ['S', 'A', 'P', ['C', 'B', 'L', 'I', 'F', ['i', 'n', 't', 'r', 'i', 'n', 's', 'i', 'c', ['f', 'a', 'c', 't', 'o', 'r',
'O', 'S', 'B', 'P', '2', ['o', 'x', 'y', 's', 't', 'e', 'r', 'i', 'o', 'l', 'b', 'i', 'n', 'd', 'i', 'n', 'g', ['2', 'O', 'S', 'B', 'P', 'r',
'4', 'o', 'x', 'y', 's', 't', 'e', 'r', 'i', 'o', 'l', ['1', 'o', 'x', 'y', 's', 't', 'e', 'r', 'i', 'o', 'l', 'b', 'i', 'n', 'd', 'i', 'n', 'g',
'n', 'r', 'e', 'l', 'a', 't', 'e', 'd', ['R', 'A', 'S', 'S', 'F', '1', '0', ['1', '0', 'R', 'a', 's', ['G', 'P', 'R', '1', '3', '2',
'c', 'c', 'u', 'm', 'u', 'l', 'a', 't', 'i', 'o', 'n', ['D', 'U', 'S', 'P', '1', '6', ['1', '6', 'M', 'A', 'P', ['7', 'M', 'A', 'P',
'a', 't', 'a', 's', 'e', '7', 'm', 'i', 't', 'o', 'g', 'e', 'n', 'a', 'c', 't', 'i', 'v', 'a', 't', 'e', 'd', ['L', 'Y', 'L', '1', ['1',
's', ['1', '8', '1', 'y', 'm', 'p', 'h', 'o', 'b', 'l', 'a', 's', 't', 'i', 'c', ['h', 'e', 'm', 'a', 't', 'o', 'p', 'o', 'i', 'e', 's',
'p', 'h', 'o', 'b', 'l', 'a', 's', 't', 'i', 'c', ['s', 'e', 'q', 'u', 'e', 'n', 'c', 'e', ['N', 'K', 'X', '2', '2', ['N', 'k', 'x',
'K', '2', ['B', 'h', 'o', 'm', 'e', 'o', 'b', 'x', ['F', 'M', 'N', '2', ['f', 'o', 'r', 'm', 'i', 'n', '2', ['L', 'A', 'T', '2',
'2', 'W', 'i', 'l', 'l', 'i', 'a', 'm', 's', 'B', 'e', 'u', 'r', 'e', 'n', ['c', 'h', 'r', 'o', 'm', 'o', 's', 'o', 'm', 'a', 'l', ['i',
'i', 'n', 'W', 'i', 'l', 'l', 'i', 'a', 'm', 's', 'B', 'e', 'u', 'r', 'e', 'n', ['p', 'r', 'o', 't', 'e', 'i', 'n', 'l', 'i', 'n', 'k',
'l', 's', 'l', 'i', 'n', 'k', 'e', 'r', ['2', 'm', 'e', 'm', 'b', 'r', 'a', 'n', 'e', 's', 's', 'o', 'c', 'i', 'a', 't', 'e', 'd',
'l', 'e', 'n', 'o', 'n', 'T', 'c', 'e', 'l', 'l', ['A', 'S', 'B', '2', ['S', 'O', 'C', 'S', ['2', 'a', 'n', 'k', 'y', 'r', 'i', 'n',
'a', 'i', 'n', 'i', 'n', 'g', ['S', 'I', 'A', 'E', ['s', 'i', 'a', 'l', 'a', 't', 'e', ['O', 'a', 'c', 'e', 't', 'y', 'l', 'e', 'i',
'L', 's', 'e', 'c', 'y', 't', 'o', 's', 'o', 'i', 'c', ['s', 'i', 'a', 'l', 'i', 'c', ['9', 'O', 'a', 'c', 'e', 't', 'y', 'l', 'e',
'e', ['h', 'm', 'o', 'l', 'o', 's', 'i', 'a', 'l', 'i', 'c', ['a', 'c', 'i', 'd', 's', 'p', 'e', 'c', 'i', 'f', 'i', 'c', ['s', 't', 'e', 'r', 'i', 'n', 'a', 's', 'e', ['M', 'L', 'X', 'I', 'P', ['M', 'L', 'X', 'i', 'n', 't', 'e', 'r', 'a', 'c', 't', 'i', 'n', 'g', ['M', 'l', 'x', ['i', 'n', 't', 'e', 'r', 'a', 'c', 't', 'o', 'l', 'c', 'l', 'a', 's', 's', ['3', '6', 't', 'r', 'a', 'n', 's', 'c', 'r', 'i', 'p', 't', ['M', 'o', 'n', 'd', 'i', 'A', ['T', 'S', 'G', 'A', '1', '0', ['t', 'e', 's', 't', 'i', 's', 'p', 'e', 'c', 'i', 'f', 'i',
'i', 'n', 'c', 'a', 'n', 'c', 'e', 'r', 't', 'e', 's', 't', 'i', 's', ['7', '9', 't', 'e', 's', 't', 'i', 's', ['N', 'Y', 'D', 'S', 'P',
's', 'p', 'e', 'c', 'i', 'f', 'i', 'c', ['D', 'P', 'H', '1', ['3', 'a', 'm', 'i', 'n', 'o', '3', 'c', 'a', 'n', 'b', 'o', 'x', 'y', 'p',
'i', 's', 't', 'i', 'd', 'i', 'n', 'e', ['1', 'D', 'P', 'H', '1', 'i', 'k', 'e', ['1', 'D', 'P', 'H', '1', ['h', 'o', 'm', 'o', 'l',
'i', 'k', 'e', ['1', 'S', 'a', 'd', 'e', 'n', 'o', 's', 'y', 'l', 'L', 'm', 'e', 't', 'h', 'i', 'o', 'n', 'i', 'n', 'e', ['L', 'h', 'i',
'e', ['3', 'a', 'm', 'i', 'n', 'o', '3', 'c', 'a', 'n', 'b', 'o', 'x', 'y', 'p', 'r', 'o', 'p', 'y', 'l', 't', 'r', 'a', 'n', 's', 'f',
'c', 'a', 'n', 'd', 'i', 'd', 'a', 't', 'e', ['1', 'd', 'i', 'p', 'h', 't', 'h', 'a', 'm', 'i', 'd', 'e', ['b', 'i', 'o', 's', 'y', 'n',
'1', 'd', 'i', 'p', 'h', 't', 'h', 'a', 'm', 'i', 'd', 'e', ['1', 'd', 'i', 'p', 't', 'h', 'e', 'r', 'i', 'a',
'm', 'i', 'd', 'e', ['S', 'a', 'c', 'c', 'h', 'a', 'r', 'o', 'm', 'y', 'c', 'e', 's', ['1', 'o', 'v', 'a', 'r', 'i', 'a', 'n', ['c',
's', 'o', 'c', 'i', 'a', 't', 'e', 'd', ['p', 'r', 'o', 't', 'e', 'i', 'n', 'o', 'v', 'a', 'r', 'i', 'a', 'n', ['T', 'P', 'D', '5', '2
'C', 'R', 'b', 'i', 'n', 'd', 'i', 'n', 'g', ['D', 'S', '2', ['U', 'H', 'K', '1', ['K', 'i', 's', 't', 'K', 'I', 'S', ['K',
'M', ['C', 'O', 'O', 'H', 't', 'e', 'r', 'm', 'i', 'n', 'a', 'l', ['i', 'n', 't', 'e', 'r', 'a', 'c', 't', 'o', 'r', ['2', 'U', '2', 'i',
'k', 'i', 'n', 'a', 's', 'e', ['s', 't', 'a', 't', 'h', 'm', 'i', 'n', ['H', 'E', 'S', '5', ['H', 'E', 'S', '5', 'c', 'l', 'a', 's',
'y', ['s', 'p', 'l', 'i', 't', ['A', 'T', 'F', '7', 'I', 'P', ['7', 'i', 'n', 't', 'e', 'r', 'a', 'c', 't', 'i', 'n', 'g', ['1', 'A',
'a', 'c', 't', 'i', 'n', 'g', ['p', 'r', 'o', 't', 'e', 'i', 'n', 'A', 'T', 'F', '7', 'i', 'n', 't', 'e', 'r', 'a', 'c', 't', 'i', 'n', 'g',
'i', 'n', 'A', 'T', 'F', 'a', 'a', 's', 's', 'o', 'c', 'i', 'a', 't', 'e', 'd', ['m', 'o', 'd', 'u', 'l', 'a', 't', 'o', 'r', 'M', 'B',
'i', 'n', 'i', 'n', 'g', ['c', 'h', 'r', 'o', 'm', 'a', 't', 'i', 'n', 'a', 's', 's', 'o', 'c', 'i', 'a', 't', 'e', 'd', ['R', 'A', 'S',
's', ['4', 't', 'u', 'm', 'o', 'r', ['B', 'L', 'I', 'D', ['B', 'H', '3', 'l', 'i', 'k', 'e', ['m', 'o', 't', 'i', 'f', 'c', 'o', 'n',
'g', ['i', 'n', 'd', 'u', 'c', 'e', 'r', 'b', 'r', 'e', 'a', 's', 't', ['B', 'C', 'L', '2', 'L', '1', '4', ['B', 'C', 'l', '2', 'l',
'L', '2', 'l', 'i', 'k', 'e', ['B', 'C', 'L', 'G', 'b', 'c', 'l', '2', 'l', '1', '4', 't', 'e', 's', 't', 'i', 'c', 'u', 'l', 'a', 'r',
'2', ['2', 'I', 'F', 'G', 'P', ['4', 'S', 'H', '2', ['a', 'n', 'c', 'h', 'o', 'r', ['1', 'f', 'c', ['2', 'i', 'm', 'm', 'u', 'n',
'o', 'g', 'l', 'o', 'b', 'u', 'l', 'i', 'n', ['g', 'p', '4', '2', ['A', 'D', 'G', 'R', 'F', '1', ['F', '1', 'G', ['1', '1', '0', 'G',
'p', 'r', 'o', 't', 'e', 'i', 'n', ['K', 'P', 'G', '1', '0', '1', '2', 'p', 'r', 'o', 'b', 'a', 'b', 'l', 'e', ['1', '1', '0', 's', 'e
'2', 'Q', '2', ['u', 'b', 'i', 'q', 'u', 'i', 't', 'i', 'n', 'c', 'o', 'n', 'j', 'u', 'g', 'a', 't', 'i', 'n', 'g', ['Q', '2', 'E', '2
'u', 'i', 't', 'i', 'n', ['c', 'o', 'n', 'j', 'u', 'g', 'a', 't', 'i', 'n', 'g', ['E', '2', 'Q', ['2', 'u', 'b', 'i', 'q', 'u', 'i',
'e', 'i', 'n', ['Q', '2', ['F', 'C', 'R', 'L', '1', ['1', 'I', 'F', 'G', 'P', ['1', 'f', 'c', 'R', 'l', 'i', 'k', 'e', ['1', 'h',
'm', 'u', 'n', 'e', ['5', 'i', 'm', 'm', 'u', 'n', 'o', 'g', 'l', 'o', 'b', 'u', 'l', 'i', 'n', ['F', 'H', 'O', 'D', '3', ['F', 'H,
'r', 'm', 'a', 'c', 't', 'i', 'n', '2', 'f', 'o', 'r', 'm', 'i', 'n', ['o', 'v', 'e', 'r', 'e', 'x', 'p', 'r', 'e', 's', 'e', 'd',
'T', 'T', 'C', '1', '2', ['1', '2', 'T', 'P', 'R', ['T', 'A', 'L', '2', ['2', 'T', 'A', 'L', '2', 'c', 'l', 'a', 's', 's', ['C', 'R
'M', 'P', 'r', 'e', 's', 'p', 'o', 'n', 's', 'i', 'v', 'e', ['2', 'M', 'H', 'B', 's', 'b', 'i', 'n', 'd', 'i', 'n', 'g', ['C', 'L', 'E
'p', 'e', ['1', 'D', 'C', 'a', 's', 's', 'o', 'c', 'i', 'a', 't', 'e', 'd', ['l', 'e', 'c', 't', 'i', 'n', '1', 'd', 'e', 'n', 'd', 'r
'l', 'l', 'a', 's', 's', 'o', 'c', 'i', 'a', 't', 'e', 'd', ['1', 't', 'y', 'p', 'e', ['D', 'C', 'A', 'L', '1', ['S', 'L', 'X', '4',
'1', ['m', 'i', 'n', 'o', 'r', ['H', 'B', '1', 'm', 'i', 'n', 'o', 'r', ['H', 'B', '1', ['Z', 'N', 'F', '2', '9', '6', ['2', '9',
'2', ['M', 'I', 'R', '5', '2', '0', 'H', ['h', 's', 'a', 'm', 'i', 'r', '5', '2', '0', 'h', ['R', 'C', 'S', 'D', '1', ['c', 'a', 'p
'c', 't', 'i', 'n', 'g', ['p', 'r', 'o', 't', 'e', 'i', 'n', 'R', 'C', 'S', 'D', ['C', 'a', 'p', 'Z', 'I', 'P', 'p', 'r', 'o', 't', 'e
'4', ['L', 'O', 'C', '1', '0', '7', '9', '8', '0', '4', '4', '0', ['r', 'e', 'c', 'o', 'm', 'b', 'i', 'n', 'a', 't', 'i', 'o', 'n', ['h',
'i', 'l', 'a', 'd', 'e', 'l', 'p', 'h', 'i', 'a', ['C', 'h', 'r', 'o', 'm', 'o', 's', 'o', 'm', 'e', ['L', 'O', 'C', '1', '0', '7
'p', '2', '1', '0', ['P', 'h', 'i', 'l', 'a', 'd', 'e', 'l', 'p', 'h', 'i', 'a', ['r', 'e', 'g', 'i', 'o', 'n', 'B', 'C', 'R', ['r',
'2', '2', ['q', '3', '4', ['q', '1', '1', ['M', 'B', 'C', 'R', ['L', 'O', 'C', '1', '0', '7', '9', '6', '3', '9', '5', '1', ['p',
'R', ['h', 'o', 'm', 'e', 'o',]]

Create a dictionary (from bigrams) and form a corpus

In [9]:

```
# Create Dictionary
id2word = corpora.Dictionary(data_words_bigrams)

# Create Corpus
texts = data_words_bigrams

# Term Document Frequency
corpus = [id2word.doc2bow(text) for text in texts]

# View
print(corpus[:1])
```

```
[(0, 1), (1, 1), (2, 1), (3, 1), (4, 1)]
```

In [10]:

```
id2word[0]
```

Out[10]:

```
'1'
```

In [11]:

```
# Human readable format of corpus (term-frequency)
[[id2word[id], freq] for id, freq in cp] for cp in corpus[:1]]
```

Out[11]:

```
[(['1', 1), ('F', 1), ('I', 1), ('K', 1), ('Z', 1)]]
```

Build LDA Model (using corpus and dictionary)

- 5 topics
- 10 passes

In [12]:

```
# Build LDA model
lda_model = gensim.models.ldamodel.LdaModel(corpus=corpus,
                                             id2word=id2word,
                                             num_topics=5,
                                             random_state=100,
                                             update_every=1,
                                             chunksize=100,
                                             passes=10,
                                             alpha='auto',
                                             per_word_topics=True)
```

In [13]:

```
# Print the Keyword in the 10 topics
print(lda_model.print_topics())
doc_lda = lda_model[corpus]
```

```
[(0, '0.248*i" + 0.165*n" + 0.137*k" + 0.097*b" + 0.095*2" + 0.083*u" + 0.055*v" + 0.048*g" + 0.036*d" + 0.029*q"'), (1, '0.226*
L" + 0.076*T" + 0.065*3" + 0.063*S" + 0.048"H" + 0.044*5" + 0.038*0"'), (2, '0.117*e" + 0.091*i" + 0.086*a" + 0.082*t" + 0.086*
2*l" + 0.061*s" + 0.052*c"'), (3, '0.177"C" + 0.155"D" + 0.148*2" + 0.098*4" + 0.098*B" + 0.069*9" + 0.068*A" + 0.064*K" + 0.0
+ 0.148*I" + 0.146*M" + 0.128*F" + 0.085"E" + 0.085*2" + 0.084*G" + 0.065*7" + 0.029"H" + 0.019*Y"')]
```

Calculate perplexity (to measure the quality of model) and calculate coherence score (more precise than perplexity)

In [14]:

```
# Compute Perplexity
print('\nPerplexity: ', lda_model.log_perplexity(corpus)) # the lower the better.

# Compute Coherence Score
coherence_model_lda = CoherenceModel(model=lda_model, texts=data_words_trigrams, dictionary=id2word, coherence='c_v')
coherence_lda = coherence_model_lda.get_coherence()
print('\nCoherence Score: ', coherence_lda)
```

```
Perplexity: -3.411499894616546
```

```
Coherence Score: 0.46729248820258135
```

Visualize topics

In [15]:

```
pyLDavis.enable_notebook()
vis = pyLDavis.gensim.prepare(lda_model, corpus, id2word)
vis
```

C:\Users\jacky\Anaconda3\lib\site-packages\pyLDavis_prepare.py:257: FutureWarning: Sorting because non-concatenation axis is not aligned of pandas will change to not sort by default.

To accept the future behavior, pass 'sort=False'.

To retain the current behavior and silence the warning, pass 'sort=True'.

```
return pd.concat([default_term_info] + list(topic_dfs))
```

Out[15]:

Selected Topic:

Previous Topic

Next Topic

Clear Topic

