

UNIVERSIDADE FEDERAL DO PARANÁ

DAVISSON HENRIQUE PAULINO, FERNANDO MEDEIROS DUFOUR

REPRESENTAÇÕES DE TEXTO, UM ESTUDO EMPÍRICO EM SISTEMAS
DE RECOMENDAÇÃO BASEADOS EM CONTEÚDO

CURITIBA PR
2018

DAVISSON HENRIQUE PAULINO, FERNANDO MEDEIROS DUFOUR

REPRESENTAÇÕES DE TEXTO, UM ESTUDO EMPÍRICO EM SISTEMAS
DE RECOMENDAÇÃO BASEADOS EM CONTEÚDO

Trabalho apresentado como requisito parcial à conclusão do Curso de Bacharelado em Ciência da Computação, setor de Ciências Exatas, da Universidade Federal do Paraná.

Área de concentração: *Ciência da Computação*.

Orientador: Eduardo Jaques Spinosa.

CURITIBA PR
2018

Resumo

A representação de documentos de texto é uma parte crítica de Sistemas de Recomendação. No entanto, para uso em tal aplicação, é necessário que as entradas de texto estejam representadas como um vetor de tamanho fixo. Neste trabalho propõe-se o uso de diferentes abordagens para representação e extração de características de textos. Analisou-se o impacto de tais representações com a criação de um Sistema de Recomendação Baseado em Conteúdo para filmes. Um resultado quantitativo foi gerado com base na coleta de uma base de avaliações de usuários online, comparando o valor real da avaliação de filmes com o predito pelo Sistema.

Palavras-chave: Extração de características, Sistemas de Recomendação, Modelagem de Tópicos, LDA, Doc2Vec, KNN.

Abstract

The representation of text documents is a critical part of Recommendation Systems. However, for use in such an application, it is necessary that the text entries be represented as a fixed-length vector. This work proposes the use of different approaches for representation and extraction of text characteristics. The impact of such representations were analyzed with the creation of a Content Based Recommendation System for films. A quantitative result was generated from an online user reviews base by comparing the actual value of the evaluation of films with the one predicted by the system.

Keywords: Feature Extraction, Recommender Systems, Topic Modelling, LDA, Doc2Vec, KNN.

Lista de Figuras

2.1	Representação em Bow dos documentos	11
2.2	Processo generativo assumido pelo LDA para geração dos documentos	11
2.3	Inferência das variáveis do modelo	12
2.4	Representação de documentos como proporção de tópicos e de tópicos como distribuição de palavras	12
2.5	A palavra <i>king</i> é para <i>queen</i> o que <i>man</i> é para <i>woman</i>	13
2.6	Exemplo CBoW.	13
2.7	Exemplo PV-DM	14
2.8	Exemplo PV-DBOW	14

Lista de Tabelas

5.1	Tópicos antes e depois da retirada de nomes comuns.	21
5.2	Tópicos gerados com K valendo 10 e 200	21
5.3	Valor do RMSD para cada cenário LDA	21
5.4	Filmes mais próximos do filme Toy Story no espaço gerado com LDA + Gênero .	22
5.5	Resultados RMSD com relação ao Doc2Vec	22

Lista de Acrônimos

AM	Aprendizado de Máquina
BoW	Bag of Words
CBow	Continuous bag of words
KNN	K-nearest neighbors
LDA	Latent Dirichlet Allocation
PV-DBOW	Distributed Bag of Words version of Paragraph Vector
PV-DM	Distributed Memory version of Paragraph Vector
RMSD	Raiz do Erro Quadrático Médio
SR	Sistemas de Recomendação

Sumário

1	Introdução	9
2	Representação de documentos de texto	10
2.1	Introdução.	10
2.2	Bag of Words	10
2.3	Modelagem de Tópicos	11
2.3.1	<i>Latent Dirichlet Allocation</i> (LDA)	11
2.4	<i>Doc2Vec</i>	12
2.4.1	Word2Vec	12
2.4.2	Extensão do Word2Vec: Doc2Vec	13
3	Sistemas de Recomendação	15
3.1	Sistemas de Recomendação Baseados em Conteúdo	15
3.2	Sistemas de Recomendação Baseados em Filtragem Colaborativa.	15
3.3	Perfil do Usuário	15
3.4	KNN.	16
4	Abordagem Proposta	17
4.1	Introdução.	17
4.2	Representação de itens	17
4.3	Aprendizagem do Perfil do Usuário	17
4.4	Recomendação de novos itens	18
5	Experimentos	19
5.1	Base de Dados.	19
5.2	Pré-processamento	19
5.3	Métricas	20
5.4	Metodologia Experimental	20
5.5	LDA	20
5.5.1	RMSD.	21
5.5.2	Recomendação	22
5.6	Doc2Vec.	22
5.6.1	RMSD.	22
5.7	Avaliando	22
6	Conclusão	23
	Referências	24

Apêndice A: Lista de stop words	25
Apêndice B: Lista nomes comuns da língua inglesa.	26

1 Introdução

A quantidade de informações geradas por dispositivos digitais tem crescido nos últimos anos com o advento das redes sociais e serviços online. Notícias, informações sobre filmes, tweets e descrições de vagas de emprego são apenas alguns exemplos da ampla gama de conteúdos existentes para consumo na internet.

Junto com essa grande oportunidade para exploração de dados surgem também desafios como a extração, representação e a devida classificação desses conteúdos que encontram-se em uma representação não estruturada.

A representação de documentos de texto é uma parte crítica de diversas aplicações de Aprendizado de Máquina (AM), como Sistemas de busca e Sistemas de Recomendação (SR). No entanto, para maioria de tais aplicações, é necessário que as entradas de texto estejam representadas como um vetor de tamanho fixo.

Um modo simples de representar documentos é através do *Bag of Words* (BoW), onde assume-se que as palavras ocorrem de forma independente. Tal modelo tem muitas desvantagens como: perda da ordem das palavras, diferentes documentos podendo ter a mesma representação e também a alta dimensionalidade/esparsidade dos vetores gerados.

Para capturar relacionamentos importantes entre as palavras, Blei et al. (2003) propôs, *Latent Dirichlet Allocation* (LDA). Tal abordagem agrupa palavras em tópicos, sendo cada documento, por sua vez, representado como uma mistura de tais tópicos.

Com o avanço de técnicas de AM, Mikolov et al. (2013) e Le e Mikolov (2014) propuseram uma representação de palavras e trechos de texto como vetores de tamanho fixo em um espaço contínuo, onde a proximidade entre os itens apresenta a similaridade semântica entre os mesmos.

Esta monografia tem como objetivo fazer uma análise sobre as diferentes abordagens utilizadas para representação de documentos de textos propondo sua utilização e mensuração de impacto em um Sistema de Recomendação para filmes baseado em conteúdo.

Este trabalho está organizado em 6 capítulos. O capítulo 2 define conceitos sobre as abordagens para representação de textos: BoW, LDA e Doc2Vec. O capítulo 3 apresenta os conceitos sobre Sistemas de Recomendação, como a diferenciação entre os Baseados em Conteúdo com relação aos de Filtragem Colaborativa. O capítulo 4 define a abordagem proposta para este trabalho. O capítulo 5 apresenta os detalhes dos experimentos realizados, como pré-processamento, assim como os resultados obtidos pelos mesmos. O capítulo 6 apresenta as considerações finais acerca do tema.

2 Representação de documentos de texto

2.1 Introdução

Ter características informativas e bem representadas é um fator crítico para o sucesso de inúmeras aplicações de Sistemas de Aprendizado de Máquina e Sistemas de Recomendação. Um espaço de características bem definido pode permitir a identificação de importantes similaridades entre os dados (Bafna e Wiens, 2015).

Dados podem ser classificados como estruturados e não estruturados. Os dados estruturados são aqueles onde as informações são encontradas de maneira rotulada e seguindo uma organização lógica, como planilhas eletrônicas e bancos de dados. Já os dados não estruturados são os encontrados de maneira heterogênea, de forma a não permitir uma previsibilidade sobre os mesmos, como é o caso de documentos de texto, imagens e vídeos.

Este capítulo tem como objetivo apresentar conceitos e métodos utilizados para extração de características de documentos de texto, de uma maneira a permitir uma melhor descrição e representação em tamanho fixo destes dados.

2.2 Bag of Words

O *Bag of Words* (BoW) é um método de extração de características simples para textos. Tal método consiste em um vocabulário formado por todas as palavras presentes no conjunto de documentos, sendo que cada palavra é representada por um inteiro único no vocabulário. Além disso, é contada a frequência de cada palavra em um documento, sendo cada documento representado por um vetor dessas frequências (Calderon, 2017). Por exemplo, considere os seguintes documentos:

1. **I love dogs.**
2. **I hate dogs and knitting.**
3. **Knitting is my hobby and my passion.**

A representação em *Bag of Words* de cada um seria dada da seguinte forma:

	I	love	dogs	hate	and	knitting	is	my	hobby	passion
Doc 1	1	1	1							
Doc 2	1		1	1	1	1				
Doc 3					1	1	1	2	1	1

Figura 2.1: Representação em Bow dos documentos

2.3 Modelagem de Tópicos

Modelagem de Tópicos é uma técnica não supervisionada que tem o objetivo de descobrir os temas principais que permeiam grandes e não estruturadas coleções de documentos. Por meio desta técnica é possível organizar e representar coleções de documentos com base nos temas descobertos (Blei et al., 2003).

Na modelagem, cada documento é representado como uma combinação de tópicos e cada tópico é representado por um conjunto de termos, ambos com probabilidades associadas. Assim, cada tópico extraído da coleção possui termos mais relevantes. Analogamente, cada documento possui tópicos mais relevantes de acordo com as respectivas probabilidades (Nolasco e Oliveira, 2016).

2.3.1 *Latent Dirichlet Allocation* (LDA)

Latent Dirichlet Allocation (LDA) é um modelo generativo probabilístico de Modelagem de Tópicos. A ideia básica é de que os documentos são representados por uma mistura aleatória de K tópicos latentes, e que cada tópico é caracterizado como uma distribuição de palavras (Blei et al., 2003).

A imagem 2.3 ilustra o processo generativo para cada documento w em um conjunto de textos D :

1. Choose $N \sim \text{Poisson}(\xi)$.
2. Choose $\theta \sim \text{Dir}(\alpha)$.
3. For each of the N words w_n :
 - (a) Choose a topic $z_n \sim \text{Multinomial}(\theta)$.
 - (b) Choose a word w_n from $p(w_n | z_n, \beta)$, a multinomial probability conditioned on the topic z_n .

Figura 2.2: Processo generativo assumido pelo LDA para geração dos documentos
Fonte: Blei et al. (2003)

O problema-chave para a criação do modelo é a inferência das variáveis implícitas, em outras palavras, é importante obter a informação da composição (θ, z) dos tópicos latentes. Dados os parâmetros da Distribuição Dirichlet α e β , as variáveis aleatórias θ , z e w no documento d são computadas como:

$$p(\theta, \mathbf{z}, \mathbf{w} | \alpha, \beta) = p(\theta | \alpha) \prod_{n=1}^N p(z_n | \theta) p(w_n | z_n, \beta),$$

Figura 2.3: Inferência das variáveis do modelo

Fonte: Blei et al. (2003)

Como resultado da inferência, cada documento pode ser representado como um vetor de proporções de tópicos e cada tópico como uma distribuição de palavras, conforme ilustra a figura 2.4.

Documentos				Tópicos			
Doc 1	1.0	0	0	Tópico 0	Carro 0.6	Moto 0.30	
Doc 2	0.333	0.333	0.333	Tópico 1	Maçã 0.4	Uva 0.2	...
Doc 3	0.88	0.1	0.02	Tópico 3	PC 0.1	Cabo 0.05	...
Doc 4	0.7	0.25	0.05				

Figura 2.4: Representação de documentos como proporção de tópicos e de tópicos como distribuição de palavras

2.4 Doc2Vec

O *Doc2Vec*, ou *Paragraph Vector*, foi proposto por Le e Mikolov (2014), sendo uma extensão do *Word2Vec*. Tal técnica permite representar documentos de tamanhos arbitrários em vetores de tamanho fixo. Para entender essa extensão primeiro será explicado como o *Word2Vec* funciona.

2.4.1 Word2Vec

O *Word2Vec* foi proposto por Mikolov et al. (2013), este método propõe resolver um dos principais problemas do BoW, a falta da representação do significado das palavras no modelo. Como no BoW as palavras são identificadas por inteiros escolhidos ao acaso, elas são desprovidas de qualquer representação semântica, ao contrário do que acontece no mundo real. O *Word2Vec* representa as palavras como vetores, permitindo descobrir o seu significado de acordo com os contextos em que aparecem, tornando possível perceber relacionamentos entre elas em um espaço gerado. Na figura 2.5 pode-se notar que a relação existente entre as palavras *man* e *woman* também existe entre as palavras *king* e *queen*.

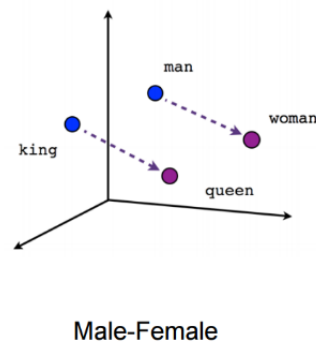


Figura 2.5: A palavra *king* é para *queen* o que *man* é para *woman*
 Fonte: Shperber (2017)

De acordo com a proposta do Word2Vec, é possível estabelecer o relacionamento entre palavras por meio de dois algoritmos. O primeiro deles é o *Skip gram*. Este algoritmo funciona posicionando uma janela em cada palavra, esta janela nada mais é do que o contexto do termo. Finalmente, uma rede neural é treinada com todo o conjunto de documentos para que dada uma palavra o modelo adivinhe as outras que devem estar na janela.

O segundo algoritmo que é passível de ser utilizado é o *Continuous bag of words* ou CBoW. Este método busca fazer o contrário do *Skip gram*. Ao invés de treinar a rede neural para descobrir o contexto de um termo, a rede é treinada para que, dado um contexto, se descubra o termo daquela janela (Shperber, 2017). A figura 2.6 mostra um exemplo do CBoW onde o contexto é formado pelas palavras “*the*”, “*cat*” e “*sat*”, sendo elas utilizadas para descobrir que o termo “*on*” é o mais provável de ter gerado tal contexto.

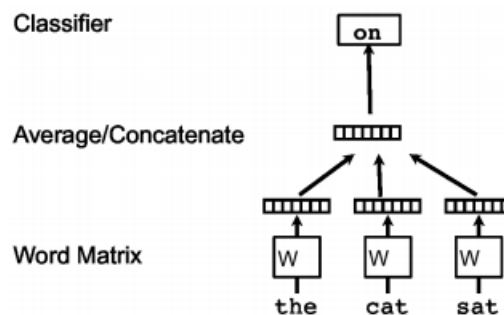


Figura 2.6: Exemplo CBoW
 Fonte: Le e Mikolov (2014)

2.4.2 Extensão do Word2Vec: Doc2Vec

Assim como no *Word2Vec*, a representação dos documentos pode ser aprendida de duas formas. A primeira é a *Distributed Memory version of Paragraph Vector* (PV-DM), ela se assemelha ao CBoW, com a diferença de que é adicionado um vetor que representa, de forma única, um documento para participar do treinamento. Ou seja, no lugar da rede neural ser treinada para, dado um termo, descobrir apenas as palavras da janela, neste modelo também será descoberto um documento. A figura 2.7 apresenta o mesmo exemplo da figura 2.6, porém é acrescentado o documento (“*Paragraph id*”) para prever o termo “*on*”.

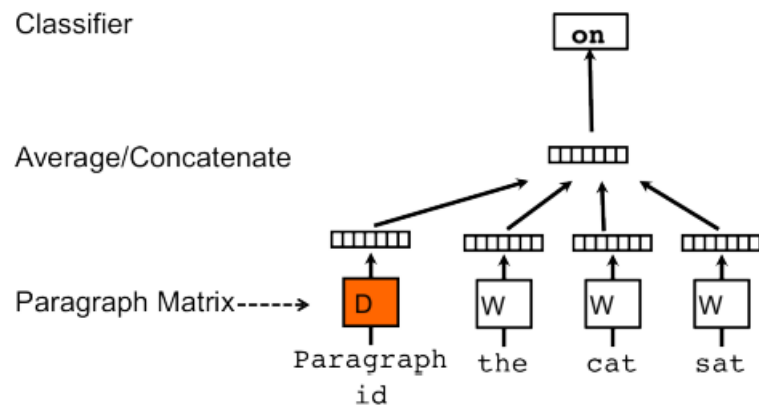


Figura 2.7: Exemplo PV-DM
Fonte: Le e Mikolov (2014)

A segunda forma é um complemento do *Skip gram* e se chama *Distributed Bag of Words version of Paragraph Vector* (PV-DBOW). Nesta forma é treinada a rede neural para que, dado um termo e o seu contexto, se infira um documento (Shperber, 2017). No exemplo da figura 2.8 a representação do documento ("Paragraph id") é descoberto a partir do termo "on" e o seu contexto.

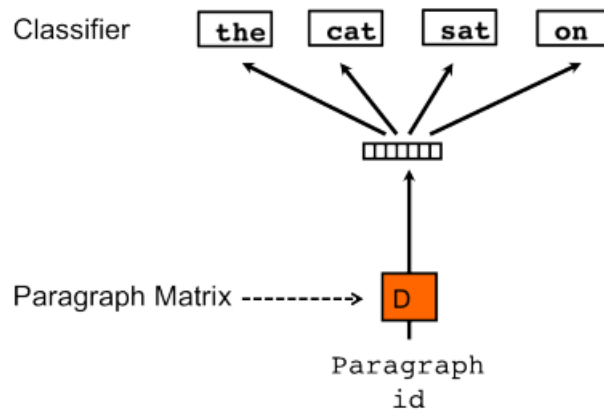


Figura 2.8: Exemplo PV-DBOW
Fonte: Le e Mikolov (2014)

A representação final de cada documento será relativa aos pesos atribuídos aos neurônios da rede neural treinada.

3 Sistemas de Recomendação

Sistemas de Recomendação (SR) são ferramentas de software e técnicas que fornecem sugestões sobre itens que possam ser úteis para um usuário. As sugestões dizem respeito a vários processos de tomada de decisão, como quais produtos comprar, qual música ouvir ou quais notícias ler.

Item é o termo geral usado para denotar o que o sistema recomenda aos usuários.

Um SR normalmente se concentra em um tipo específico de item (por exemplo, CDs, filmes ou notícias) (Ricci et al., 2010).

Abaixo serão apresentados os Sistemas de Recomendação Baseados em Conteúdo e os Sistemas de Recomendação Baseados em Filtragem Colaborativa.

3.1 Sistemas de Recomendação Baseados em Conteúdo

Neste tipo de SR, o sistema aprende a recomendar itens similares aos que o usuário gostou no passado. A similaridade é calculada com base nos Vetores de Características associados aos itens comparados. Por exemplo, se o usuário recomendou positivamente um filme do gênero comédia, então o sistema pode aprender a recomendar outros filmes que pertencem a esse gênero (Ricci et al., 2010).

3.2 Sistemas de Recomendação Baseados em Filtragem Colaborativa

Segundo Ricci et al. (2010) esta abordagem de SR recomenda para o usuário ativo itens que outros usuários com gostos similares gostaram no passado. A similaridade nos gostos de dois usuários é calculada com base na similaridade dos itens avaliados por ambos.

3.3 Perfil do Usuário

O Perfil do Usuário no contexto de Sistemas de Recomendação é o histórico de todos os itens que o usuário avaliou, seja de maneira explícita ou implícita. Por exemplo, em um sistema de recomendação de filmes pode-se compreender o Perfil como sendo a coleção de todos os filmes que o usuário associou uma nota.

3.4 KNN

O KNN ou K-vizinhos mais próximos é um método que pode ser usado tanto para uma tarefa de classificação quanto para regressão, nesse trabalho ele será experimentado no contexto de regressão.

O método recebe como entrada vetores de um espaço, rótulos associados a cada um deles e um número inteiro K . O objetivo é, para um novo vetor de entrada T , descobrir qual será seu rótulo. Esta tarefa é realizada encontrando os K -vetores mais próximos de T , desta forma, define-se o rótulo de T como a média dos rótulos dos K -vizinhos.

É possível fornecer outros parâmetros para este modelo, como: métrica de similaridade a ser usada e uma função substituta ao cálculo da média, como por exemplo uma média ponderada pela distância.

4 Abordagem Proposta

4.1 Introdução

Um dos fatores essenciais para o bom funcionamento de Sistemas de Recomendação Baseados em Conteúdo é a devida representação dos itens do Sistema. No contexto de sistemas de recomendação de filmes, informações como gênero, atores, e diretor são comumente utilizadas para tal tarefa, através de técnicas como a binarização. Por outro lado, variáveis como sinopse, enredo e críticas, por serem dados não estruturados e encontrados com variadas dimensões, demandam outras técnicas de processamento.

Tendo este cenário em vista, propõe-se o uso de diferentes abordagens para representação e extração de características aplicadas ao enredo de filmes. Analisou-se o impacto de cada abordagem com a criação de um Sistema de Recomendação Baseado em Conteúdo, avaliando tanto a fase de aprendizado do Perfil do Usuário como a recomendação em si.

4.2 Representação de itens

Devido ao enredo nem sempre carregar todas as informações contextuais de um filme, optou-se por representar cada Item do Sistema como o resultado da concatenação da informação binarizada dos gêneros com as seguintes representações, a depender do método escolhido:

- LDA: Vetor de proporções de tópicos de um documento.
- Doc2Vec: Vetor de característica descrevendo o contexto de um documento.
 - Utiliza-se o algoritmo PV-DM para gerar o vetor pois este proporciona um melhor desempenho (Le e Mikolov, 2014).

4.3 Aprendizagem do Perfil do Usuário

Assim como na abordagem proposta por Luostarinen e Kohonen (2013), implementou-se um Sistema de Recomendação de maneira isolada para cada usuário. É possível resumir as etapas envolvidas no processo de aprendizado da seguinte forma:

1. Para cada usuário, coletou-se um conjunto de filmes aos quais o usuário associou notas com o valor entre 0 e 5. Exemplos de notas seriam: 5; 1,5; 3,5; 4.
2. Para este conjunto, criou-se a representação de cada Item no formato de vetor de características, como citado na Seção 4.2.
3. Finalmente, o conjunto destes vetores de características são fornecidos para um modelo KNN de regressão, juntamente com suas notas associadas.

4.4 Recomendação de novos itens

Levando em consideração que a recomendação será feita por um SR Baseado em Conteúdo, é desejado que o sistema de recomendação de filmes consiga encontrar, dentre o universo completo de itens, aqueles mais próximos, ou seja, mais similares, aos quais o usuário informou as maiores notas. Utilizou-se a métrica de similaridade do cosseno neste trabalho, conforme apresentado posteriormente na seção 5.3.

A recomendação de novos itens foi realizada da seguinte forma:

1. Para cada filme que o usuário ainda não avaliou:
 - (a) É fornecido para o KNN a representação vetorial deste
 - (b) O KNN infere a nota que o usuário daria para aquele filme com base nos K filmes mais similares já avaliados pelo usuário
2. Uma lista é gerada, estando ordenada da maior até a menor nota predita pelo modelo.
3. Os itens podem ser recomendados de acordo com as notas inferidas

5 Experimentos

5.1 Base de Dados

Foram utilizadas duas bases de dados de catálogos de filme disponíveis online. MovieLens 1M, criada por Harper e Konstan (2015), composta por 20 milhões de avaliações de usuários aplicadas a 27.000 filmes por 138.000 usuários e CMU Movie Summary Corpus, criada por Bamman et al. (2013), composta por enredos extraídos do Wikipédia de 42.306 filmes. Ambas as bases têm seus conteúdos na língua inglesa.

Com o objetivo de gerar uma base única, realizou-se a junção das bases usando o título do filme como chave. Tal processo resultou em um total de 10.406 filmes com informações de gêneros e enredo.

Também foram filtrados somente os usuários com mais de 50 avaliações de filmes para o aprendizado do perfil e para recomendação, tais avaliações foram também divididas, aleatoriamente, em duas amostras, uma de treino, contendo 80% do total, e outra de teste, contendo 20%.

5.2 Pré-processamento

Para melhorar a representação do enredo, foram removidas palavras consideradas de baixa relevância para o contexto de recomendação, além de agregar palavras com o mesmo radical. As etapas de pré-processamento foram as seguintes:

- Remoção de palavras irrelevantes para identificar um filme (Stop words). Lista no Apêndice A
- Remoção de nomes comuns da língua inglesa. Lista no Apêndice B
- Lematização - substituição das flexões de palavras para um radical em comum.
 - Por exemplo, as palavras “am”, “are” e “is” serão todas substituídas pela palavra “be”.
- Remoção de palavras que aparecem em menos do que 10 enredos ou mais que 50% dos documentos.
- Representação de bigramas (duplas de palavras que equivalem a uma única entidade) como uma única palavra.
 - Por exemplo, “Police Station” representa uma única entidade.

5.3 Métricas

Foram utilizadas duas métricas no sistema, a Similaridade do Cosseno e a Raiz do Erro Quadrático Médio (RMSE). A primeira foi utilizada para calcular a proximidade de um filme de outro no espaço gerado. Então sendo θ o ângulo formado entre os vetores A e B de dois filmes, a similaridade é calculada da seguinte forma:

$$\cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} \quad (5.1)$$

Já a RMSE serviu para validar os experimentos, comparando as notas dadas pelos usuários a um filme com as que o sistema inferiu com base nas notas anteriores. Esta medida mostra a diferença média entre as notas e além disso penaliza erros grandes, ao contrario do Erro Quadrático Médio (Chai e Draxler, 2014). Sendo n o número de de usuários, m os filmes do conjunto de testes que foram avaliados pelo usuário i , y_{ij} a nota dada pelo usuário i no filme j e $\hat{y}_{ij}$ a nota inferida pelo sistema no mesmo, a Raiz do Erro Quadrático Médio é calculada da seguinte maneira:

$$RMSE = \sqrt{\frac{1}{n * m} \sum_{i=1}^n \sum_{j=1}^m (y_{ij} - \hat{y}_{ij})^2} \quad (5.2)$$

5.4 Metodologia Experimental

O sistema foi feito em Python3, utilizando para o pré-processamento biblioteca spacy em conjunto com a nltk; para treino dos modelos LDA e Doc2Vec utilizou-se a biblioteca Gensim e para regressão com o KNN utilizou-se o scikit-learn.

Para avaliar o impacto dos diferentes modelos de representação de texto e também os parâmetros relativos a cada um, foram realizados diversos experimentos. Para o modelo LDA, o experimento consistia em descobrir qual era o número de tópicos K que melhor representava o conjunto de enredos de filmes. Para o modelo Doc2Vec variou-se o numero de características do vetores dos documentos .

Para o aprendizado do perfil do usuário e recomendação, utilizou-se o modelo KNN para regressão, como citado no Capítulo 3. Para este modelo — adotando a quantidade de avaliações do usuário de treino como sendo N — os parâmetros utilizados foram:

- Quantidade de vizinhos analisados: $\sqrt{N}$;
- Peso para os vizinhos: Inversamente proporcional à distância;
- Métrica de Similaridade: Cosseno.

5.5 LDA

Inicialmente realizou-se uma análise sobre o impacto da remoção de nomes comuns da língua inglesa na representação dos tópicos. Tal procedimento resultou em tópicos com descrição aumentada, a tabela 5.1 abaixo exemplifica esta mudança, assume-se X como sendo um tópico antes do processo e Y um tópico após.

Tópico X	Tópico Y
david	army
bond	kill
taylor	battle
terry	man
carl	fight
melanie	soldier

Tabela 5.1: Tópicos antes e depois da retirada de nomes comuns

Realizou-se uma análise qualitativa sobre os tópicos extraídos pelo LDA conforme a variação do parâmetro de número de tópicos do modelo, referenciado como K. Notou-se que conforme o número de tópicos é aumentado, o nível de especialização das palavras presentes em cada uma destas representações também aumenta.

A tabela 5.2 abaixo mostra exemplos de tópicos, com os 5 termos mais relevantes associados a cada um, ao assumir K valendo 10 e 200.

LDA K=10			LDA K=200		
Tópico 7	Tópico 8	Tópico 4	Tópico 41	Tópico 42	Tópico 56
game	soldier	life	drug	car	british
team	army	money	deal	drive	military
play	war	family	drug_dealer	accident	officer
win	man	film	use	crash	american
school	battle	give	heroin	driver	nazi

Tabela 5.2: Tópicos gerados com K valendo 10 e 200

Nota-se que com K valendo 10, os tópicos gerados assemelham-se a gêneros de filmes, já com K valendo 200 assemelham-se a temas mais específicos.

5.5.1 RMSD

Para realização do cálculo desta métrica foram utilizados os valores de K iguais a [10, 50, 100, 150]. Além disso, para cada K realizou-se a variação da forma de representação dos filmes de duas formas distintas. No primeiro cenário, referenciado como ‘LDA + Gêneros’, utilizou-se o vetor de proporções de tópicos concatenado com a representação binarizada dos gêneros. Já no segundo cenário, utilizou-se somente o vetor de proporções de tópicos, sendo este cenário referenciado como somente ‘LDA’. A tabela 5.3

	K=10	K=50	K=100	150
LDA + Gêneros	0,9604	0,9534	0,9537	0,9556
LDA	0,9802	0,9677	0,9652	0,9673
Gêneros	1.0088			

Tabela 5.3: Valor do RMSD para cada cenário LDA

Nota-se a melhor performance da predição conforme o aumento do número de tópicos K, assim como apenas uma pequena diferença ao agregar a informação dos gêneros. Tal fenômeno deve-se ao fato de que com o número de tópicos sendo maior, o vetor resultante do LDA consegue assemelhar-se à representação do gênero dos filmes.

5.5.2 Recomendação

Analisou-se qualitativamente as recomendações geradas concluindo-se que a representação ‘LDA + Gêneros’ trouxe melhores resultados. A Tabela 5.4 demonstra a diversificação de gêneros resultante da adoção do vetor de proporções do LDA.

Filme	Gêneros
Toy Story 3	Adventure; Animation; Children; Comedy; Fantasy; IMAX
Toy Story 2	Adventure; Animation; Children; Comedy; Fantasy
Monsters; Inc.	Adventure; Animation; Children; Comedy; Fantasy
Space Jam	Adventure; Animation; Children; Comedy; Fantasy; Sci-Fi
Robots	Adventure; Animation; Children; Comedy; Fantasy; Sci-Fi; IMAX
Shrek	Adventure; Animation; Children; Comedy; Fantasy; Romance
Brother Bear 2	Adventure; Animation; Children; Comedy; Fantasy
Enchanted	Adventure; Animation; Children; Comedy; Fantasy; Musical; Romance

Tabela 5.4: Filmes mais próximos do filme Toy Story no espaço gerado com LDA + Gênero

5.6 Doc2Vec

5.6.1 RMSD

Para realização do cálculo desta métrica foi-se variado o parâmetro S , referente ao tamanho da janela de contexto utilizado no treinamento do modelo no formato . Foram-se utilizados os valores de S iguais a [20, 50, 100, 150]. Além disso, para cada S realizou-se a variação da forma de representação dos filmes de duas formas distintas. No primeiro cenário, referenciado como ‘Doc2Vec + Gêneros’, utilizou-se o vetor do documento concatenado com a representação binarizada dos gêneros. Já no segundo cenário, utilizou-se somente o vetor do documento gerado pelo *Doc2Vec*, sendo este cenário referenciado como somente ‘Doc2Vec’.

	S=20	S=50	S=100	S=150
Doc2Vec + Gêneros	0.9571	0.9581	0.9574	0.9573
Doc2Vec	0.9799	0.9821	0.9812	0.9815
Gêneros	1.0088			

Tabela 5.5: Resultados RMSD com relação ao Doc2Vec

5.7 Avaliando

Com resultados dos experimentos nota-se que o enredo pode ser um componente relevante para melhorar a recomendação de filmes. E que não há grandes diferenças entre a representação do enredo como um vetor de mistura de tópicos, utilizando LDA, ou como um vetor gerado pelo *Doc2Vec*. Além disso, os resultados do sistema como um todo se mostram similares aos trabalhos de viés próximos encontrados: Kose et al. (2017) e Bhowmick et al. (2014).

6 Conclusão

Este trabalho apresentou diversos modelos para representação de textos, assim como as melhorias que cada um propõe.

Foi escolhido o cenário de recomendação de filmes para testar os modelos. Se obteve dados sobre avaliações feitas por usuários, enredo e gêneros de filmes a partir de duas bases, Movielens 20M e CMU Movie Summary Corpus. Então, foram treinados dois modelos para a representação dos enredos, o LDA, que encontrou tópicos escondidos, atribuindo a cada enredo um vetor de porcentagens dos tópicos encontrados; e o Doc2Vec, que gerou um vetor de características para cada enredo de acordo com o seu contexto.

Representando cada filme como vetor formado pelo enredo, treinado com LDA, mais o gênero notou-se que dividir os enredos em 50 tópicos produzem bons resultados. E representando cada filme da mesma forma, porém treinando o enredo com Doc2Vec, notou-se que um vetor de 20 características também produzem resultados semelhantes, sendo o desempenho dos dois modelos similares de forma geral no caso teste deste trabalho.

Para avaliar o desempenho do sistema foi utilizada a métrica da Raiz do Erro Quadrático Médio, comparando as notas de fato dadas pelos usuários com notas inferidas pelo sistema. A inferência das notas foi feita com uma regressão através do método KNN, sendo o desempenho do sistema como um todo semelhante aos trabalhos próximos encontrados.

Referências

- Bafna, A. e Wiens, J. (2015). Automated feature learning: Mining unstructured data for useful abstractions. Em *2015 IEEE International Conference on Data Mining*. IEEE.
- Bamman, D., O'Connor, B. e Smith, N. (2013). Cmu movie summary corpus. <http://www.cs.cmu.edu/~ark/personas/>. Acessado em 06/07/2018.
- Bhowmick, A., Prasad, U. e Kottur, S. (2014). Movie recommendation based on collaborative topic modeling.
- Blei, D. M., Ng, A. Y. e Jordan, M. I. (2003). Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022.
- Calderon, P. (2017). Bag of words and tf-idf explained. <http://datameetsmedia.com/bag-of-words-tf-idf-explained/smedia.com/bag-of-words-tf-idf-explained>. Acessado em 06/07/2018.
- Chai, T. e Draxler, R. R. (2014). Root mean square error (rmse) or mean absolute error (mae)? “arguments against avoiding rmse in the literature. *Geoscientific Model Development*, 7(3):1247–1250.
- Harper, F. M. e Konstan, J. A. (2015). The movielens datasets: History and context. *ACM Trans. Interact. Intell. Syst.*, 5(4):19:1–19:19.
- Kose, A., Kanbak, C. e Evirgen, N. (2017). Performance comparison of algorithms for movie rating estimation. *CoRR*, abs/1711.01647.
- Le, Q. V. e Mikolov, T. (2014). Distributed representations of sentences and documents. *CoRR*, abs/1405.4053.
- Luostarinen, T. e Kohonen, O. (2013). Using topic models in content-based news recommender systems. Em *NODALIDA*.
- Mikolov, T., Chen, K., Corrado, G. e Dean, J. (2013). Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781.
- Nolasco, D. e Oliveira, J. (2016). Modelagem de tópicos e criação de rótulos: Identificando temas em dados semi-estruturados e não-estruturados. *Simpósio Brasileiro de Bancos de Dados (SBBD 2016)*.
- Ricci, F., Rokach, L., Shapira, B. e Kantor, P. B. (2010). *Recommender Systems Handbook*. Springer-Verlag, Berlin, Heidelberg, 1st edition.
- Shperber, G. (2017). A gentle introduction to doc2vec. <https://medium.com/scaleabout/a-gentle-introduction-to-doc2vec-db3e8c0cce5e>. Acessado em 06/07/2018.

Apêndice A: Lista de stop words

i, me, my, myself, we, our, ours, ourselves, you, you're, you've, you'll, you'd, your, yours, yourself, yourselves, he, him, his, himself, she, she's, her, hers, herself, it, it's, its, itself, they, them, their, theirs, themselves, what, which, who, whom, this, that, that'll, these, those, am, is, are, was, were, be, been, being, have, has, had, having, do, does, did, doing, a, an, the, and, but, if, or, because, as, until, while, of, at, by, for, with, about, against, between, into, through, during, before, after, above, below, to, from, up, down, in, out, on, off, over, under, again, further, then, once, here, there, when, where, why, how, all, any, both, each, few, more, most, other, some, such, no, nor, not, only, own, same, so, than, too, very, s, t, can, will, just, don, don't, should, should've, now, d, ll, m, o, re, ve, y, ain, aren, aren't, couldn, couldn't, didn, didn't, doesn, doesn't, hadn, hadn't, hasn, hasn't, haven, haven't, isn, isn't, ma, mightn, mightn't, mustn, mustn't, needn, needn't, shan, shan't, shouldn, shouldn't, wasn, wasn't, weren, weren't, won, won't, wouldn, wouldn't

Apêndice B: Lista nomes comuns da língua inglesa

0, 1, 2, 3, 4, 5, 6, 7, 8, 9, a, aaron, abbey, abbie, abdul, abe, across, abel, abigail, about, above, abraham, abram, abst, accordance, according, act, actually, ada, adah, adalberto, adaline, adam, adan, added, among, addie, adela, adelaida, adelaide, adele, adelia, adelina, adeline, adell, adella, adelle, adena, adina, adj, adolfo, adolph, adopted, adria, adrian, adriana, adriane, adrianna, adrien, adrienne, after, afterwards, afton, again, against, agatha, agnes, agnus, agueda, agustina, ahmad, ahmed, ai, aida, besides, aide, aiko, aileen, ailene, aimee, aja, akilah, al, alaina, alaine, alan, alana, alane, alanna, alayna, alba, albert, alberta, albertha, albertina, albertine, alberto, albina, alda, alden, aldo, alease, alec, alecia, e, aleen, aleisha, eg, alejandra, alejandrina, alejandro, alena, elsewhere, alene, alesha, aleshia, alesia, alessandra, aleta, aletha, everywhere, alethea, alethia, alex, alexander, alexandria, alexia, alexis, alfonso, alfonzo, alfred, alfreda, alfredia, alfredo, ali, alia, alica, alice, alicia, alida, alina, alisa, alise, had, alisha, alishia, alisia, alison, alissa, alita, alix, aliza, all, alla, allan, alleen, allegra, allen, allena, allene, allie, alline, allison, allyn, allyson, alma, hop, almeda, almeta, almost, alona, alone, along, alonso, alonzo, alpha, alphonse, alphonso, already, also, alta, altagracia, altha, althea, although, alton, alva, alvaro, alvera, alverta, alvin, alvina, always, alyce, alycia, alysa, alyse, alysha, alysia, alyson, alyssa, am, amado, amal, amalia, amanda, amber, amberly, ambrose, amee, amelia, long, america, m, ami, amie, amiee, amina, may, amira, ammie, amongst, amos, might, amparo, amy, an, ana, anabel, anamaria, anastacia, anastasia, and, andera, anderson, andra, andre, na, andrea, andreas, andres, andrew, andria, andy, anette, angel, angela, angele, angelena, angeles, angelia, angelic, angelica, angelina, angeline, angelique, angelita, angella, angelo, angelyn, angie, angila, angla, angle, anglea, anh, anibal, anika, anisha, anissa, anita, anitra, anjanette, anjelica, ann, anna, annabel, annabell, annabelle, overall, annalee, annalisa, annamae, annamaria, annamarie, anne, anneliese, annelle, annemarie, put, annetta, annice, annie, rather, annika, annis, annmarie, announce, another, answered, anthony, antione, antionette, antoine, anton, antone, antonetta, antonette, antonia, antonietta, antonina, antonio, antony, antwan, any, anya, anyhow, anyone, anything, anywhere, apolonia, april, apryl, ara, should, araceli, aracelis, aracely, arcelia, archie, ardash, ardelia, ardell, ardella, ardelle, arden, ardis, ardith, are, aren, arent, aretha, argelia, argentina, ariana, arianna, arianne, arica, arie, ariel, arielle, arla, arlean, arleen, arlen, there, arlena, arlene, arletha, arletta, arlette, arlie, arlinda, arline, arlyne, armand, armanda, they, armandina, armando, armida, arminda, arnetta, arnette, arnita, arnold, arnoldo, arnulfo, around, arron, three, art, arthur, artie, arturo, arvilla, as, asa, asha, ashanti, ashely, ashlea, toward, ashlee, ashleigh, ashley, ashli, ashlie, ashly, ashlyn, asia, ask, unlikely, asked, asley, assunta, astrid, ups, at, athena, aubrey, audie, audra, audrea, audrey, audria, audrie, audry, august, augusta, augustina, augustine, augustus, aundrea, aura, aurea, aurelia, aurelio, aurora, aurore, austin, auth, autumn, ava, available, avelina, avery, avis, avril, awilda, ayako, ayana, ayanna, ayesha, azalee, azucena, azzie, b, babara, babette, back, bailey, bambi, bao, barabara, barb, barbara, barbera, barbie, barbra, bari, barney, barrett, barrie, bart, basil, would, basilia, be, bea, beata, beatrice, beatris, beatriz, beau, beulah, bebe, became, because, becki, becky, become, becomes, becoming, bee, abby, been, before, beforehand, begin, beginning, behind, being, belen, belia, belinda, belkis, bell, bella, belle, below, ben, benedict, benita, benito, benjamin, bennett, benny, benton, berenice, berna, bernadette, bernadine, bernard,

bernarda, bernardina, bernardine, bernardo, adrienne, berneice, adriene, bernice, bernie, berniece, bernita, bert, agripina, berta, agustin, berth, bertie, bertram, beryl, beside, bess, bessie, best, beth, bethanie, aisha, bethann, akiko, bethany, bethel, betsey, bette, better, bettie, bettina, betty, bettyann, bettye, between, beula, beulah, bev, beverlee, beverley, beverly, beyond, bianca, bibi, bill, billie, billy, billye, aleida, birdie, birgit, blaine, blair, blake, blanca, blanch, blanche, blondell, blossom, blythe, bo, bobbi, bobbie, bobbie, alexa, bok, alexandra, bong, bonita, bonnie, bonny, booker, boris, both, boyce, boyd, brad, bradford, bradley, bradly, brady, brain, branda, aline, brande, brandee, branden, brandi, brandon, brandy, brant, breana, breann, breanna, breanne, bree, brenda, brendan, brendon, brenna, brent, brenton, bret, brett, brian, briana, brianne, brice, bridget, bridgett, bridgette, brigette, brigid, brigida, brigitte, brinda, britany, britney, britni, britt, britta, brittaney, brittani, brittanie, brittney, brittney, brock, broderick, bronwyn, brook, brooke, brooks, bruce, amada, bruna, brunilda, bruno, bryan, bryanna, bryant, bryce, brynn, bryon, bud, buddy, buena, buffy, buford, bula, bunny, burl, burma, burt, burton, buster, but, analisa, by, byron, c, ca, caitlin, caitlyn, calandra, caleb, calista, andree, callie, calvin, camelia, camellia, cameron, camie, camila, camilla, camille, cammy, can, candace, candance, angelika, candelaria, candi, candice, candida, candis, candy, candyce, cannot, cant, caprice, caption, cara, caren, carey, cari, anisa, caridad, carie, carina, carisa, anja, carissa, carita, carl, carla, carlee, carleen, carlena, carlene, carletta, carley, carli, carline, carlita, carlo, carlos, carlota, annett, carlotta, annette, carlton, carly, carlyn, carma, annita, carman, carmel, carmela, carmelia, carmelina, antoinette, carmelita, carmella, carmelo, carmen, carmina, carmine, carmon, carol, carola, carolann, carole, carolee, carolin, caroline, caroll, carolyn, carolyne, carolynn, caron, carolyn, carri, carrie, carrol, carry, carson, cary, caryl, carylon, caryn, casandra, casey, casie, ariane, casimira, cassandra, cassaundra, cassey, cassidy, cassie, cassondra, cassy, catalina, catarina, caterina, catharine, catherin, catherina, catherine, cathern, catheryn, catheryn, cathie, cathie, cathleen, cathrine, cathryn, cathy, catina, catrice, catrina, cayla, cecil, cecila, cecile, aron, cecilia, cecille, cecily, cedric, cedrick, celena, celesta, celeste, celestina, celestine, celia, celina, celinda, celine, celsa, ceola, cesar, chad, ashton, chadwick, chae, chan, chana, asuncion, chance, chanda, chandra, chanel, chanell, chanelle, chang, chantal, chantay, chante, chantel, chantell, chantelle, chapter, chara, charis, charise, charissa, charisse, charita, charity, charla, charlena, charlene, charles, charlesetta, charlette, charley, charlie, charline, charlott, charlotte, charlsie, charlyn, charmaine, charolette, chase, chasidy, chasity, chastity, chau, chauncey, chaya, barbar, chelsea, chelsey, chelsie, cher, chere, cheree, cherelle, cheri, barry, cherie, barton, cherilyn, cherish, cherlyn, cherri, cherrie, cherry, cherryl, chery, cheryl, cheryle, cheryll, bekie, chester, chet, cheyenne, chi, chia, chieko, chin, china, ching, belva, chiquita, chloe, chong, chris, chrissy, christa, bennie, christal, christeen, christel, christen, christene, christi, christia, christian, christiana, christiane, christie, christin, bernetta, christina, christine, christinia, christoper, berry, christopher, christy, chrystal, chu, chuck, chun, chung, cicely, ciera, cierra, cinda, cinderella, cindi, cindy, betsy, cinthia, cira, clair, clara, clare, clarence, claretha, claretta, claribel, clarice, clarine, claris, clarisa, clarissa, clarita, billi, clark, classie, claud, claude, claudette, claudia, claudie, claudine, claudio, clay, clayton, clemencia, clement, clemente, clementina, bob, clementine, clemmie, bobby, cleo, bobette, cleopatra, cleora, cleotilde, cleta, cletus, cleveland, cliff, clifford, clifton, clint, clinton, clora, clorinda, clotilde, clyde, co, codi, cody, colby, coleen, brandie, coleman, colene, coletta, colette, colin, colleen, collen, collene, collette, collin, columbus, come, concepcion, conception, concetta, concha, conchita, connie, brianna, conrad, constance, consuela, consuelo, contessa, cora, coral, coralee, coralie, corazon, cordelia, cordell, cordia, cordie, coreen, corene, coretta, corey, brittany, corie, brittney, corina, corine, corinna, corinne, corliss, cornelia, cornelius, cornell, corrie, corrin, corrina, corrine, corrinne, cortez, cortney, could, couldnt, courtney, buck, coy, craig, creola, cried, cris, criselda, bulah, crissy, crista, cristal, cristen, cristi, cristie, cristin, cristina, cristine,

cristobal, cristopher, cristy, cruz, crysta, crystal, crystle, cuc, cami, curt, curtis, cyndi, cyndy, cammie, cynthia, cyril, cyrstal, cyrus, cythia, d, dacia, candie, dagmar, candra, dagny, dahlia, daina, daine, daisey, daisy, dakota, dale, dalene, carin, dalia, dalila, dallas, dalton, damaris, damian, damien, damion, damon, dan, dana, danae, dane, carlie, danelle, danette, dani, danial, danica, daniel, daniela, daniele, daniell, daniella, danielle, danika, danille, danilo, danita, dann, danna, dannette, dannie, dannielle, danny, danuta, danyel, danyell, danyelle, daphine, dara, darby, carolina, darcel, darcey, darci, darcie, darcy, darell, daren, daria, darin, dario, carroll, darius, darla, carter, darleen, darlena, darlene, darline, darnell, daron, darrel, darrell, darren, darrick, darrin, cassi, darron, darryl, darwin, daryl, date, dave, david, davida, davina, davis, dawn, dawna, dawne, dayle, dayna, daysi, deadra, dean, deana, deandra, deandre, deandrea, deane, deangelo, cecelia, deann, deanna, deanne, deb, debbi, debbie, debbra, debby, debera, debi, debora, deborah, debra, debrah, debroah, dede, dedra, dee, deeann, deeanna, deedee, deena, deetta, deidra, deidre, deirdre, deja, del, delana, delbert, delcie, delena, delfina, delia, delicia, delila, delilah, delinda, delisa, dell, della, delma, delmar, delmer, delmy, delois, charleen, deloise, delora, deloras, delores, deloris, delorse, delpha, delphia, delphine, delsie, delta, demarcus, charmain, demetra, demetria, chas, demetrice, demetrius, dena, chassidy, denae, deneen, denese, denice, denis, denise, denisha, denita, denna, dennis, dennise, denny, denver, denyse, cherise, deon, cherly, deonna, derek, derick, derrick, deshawn, desirae, desire, desiree, despina, dessie, destiny, detra, devin, devon, devona, devora, devorah, dewayne, dewey, dewitt, dexter, dia, diamond, dian, diana, diane, diann, dianna, christena, dianne, dick, did, didnt, diedra, diedre, diego, dierdre, digna, dillon, dimple, dina, dinah, dino, dinorah, dion, dione, dionna, dionne, ciara, dirk, divina, dixie, do, dodie, does, cindie, doesnt, dollie, dolly, dolores, claire, doloris, domenic, domenica, dominga, domingo, dominic, dominica, clarinda, dominick, dominique, dominque, domitila, domonique, don, dona, donald, donella, donetta, donette, dong, donita, donn, donna, donnell, clelia, donnetta, donnette, donnie, donny, donovan, dont, donte, donya, dora, dorathy, dorcas, doreatha, doreen, dorene, doretha, dorethea, doretta, dori, doria, dorian, dorie, dorinda, dorine, doris, dorla, cole, dorothea, dorothea, dorothy, dorris, dorsey, dortha, dortha, dorthey, dorthy, dot, dotty, colton, doug, douglas, douglass, dovie, down, doyle, dreama, drew, drucilla, duane, dudley, dulcie, dung, during, dusti, dustin, dusty, dwana, dwayne, dwight, dylan, each, earl, earle, earlean, cori, earleen, earlene, earlie, earline, earnest, earnestine, eartha, easter, eboni, ebonie, ebony, echo, ed, edda, eddie, cory, eddy, edelmira, eden, edgardo, edie, edith, edmond, edmund, edmundo, edna, edra, edris, eduardo, edward, eduardo, edwin, edyth, edythe, effie, efrain, efren, ehtel, eight, eighty, eilene, either, ela, eladia, elaina, elaine, elana, elane, elanor, elayne, elba, elda, elden, eldon, eldora, eldridge, eleanor, eleanora, eleanore, elease, elena, elene, eleni, elenor, elenora, eleonor, eleonora, eleonore, elfreda, elfrieda, elfriede, eli, elia, eliana, elias, dania, elicia, elida, elidia, elijah, elina, elinor, elinore, elisa, elisabeth, elise, eliseo, elisha, elissa, eliz, eliza, elizabeth, elizabeth, elizabeth, dante, elke, ella, ellamae, ellan, ellen, daphne, ellena, elli, ellie, elliot, elliot, ellis, ellsworth, elly, ellyn, elma, elmer, elmira, elmo, elna, elnora, elodia, elois, elois, eloise, elouise, eloy, elroy, elsa, else, elsie, elsy, elton, elva, elvera, elvia, elvie, elvina, elvira, elvis, elwanda, elwood, elyse, elza, ema, emanuel, emelda, emelia, emelina, emeline, emely, emerald, emerita, emerson, emery, emiko, emil, emile, emilee, emilia, emilie, emilio, emily, emma, emmaline, emmanuel, emmie, emmitt, emmy, emogene, emory, ena, enda, enedina, deetra, enid, enoch, enola, enough, enrique, enriqueta, epifania, delaine, era, erasmo, eric, erica, erich, erick, ericka, erik, erika, erin, erinn, erlene, erlinda, erline, erma, ermelinda, erminia, erna, ernest, ernestina, ernestine, ernesto, ernie, errol, erwin, eryl, esmeralda, esperanza, essie, esteban, estefana, estela, estell, estella, estelle, ester, esther, estrella, etc, etha, ethan, denisse, ethel, ethelene, ethelyn, ethyl, etsuko, etta, ettie, eufemia, eugena, eugene, eugenia, eugenie, eugenio, eula, eulah, eulalia, desmond, eun, euna, eunice, eura, eusebia, eusebio, eustolia, evalyn, evan,

evangelina, evangeline, eve, evelia, evelin, evelina, eveline, evelyn, evelyne, evelynn, even, ever, everett, everette, every, everyone, everything, evette, evia, evie, evita, evon, evonne, ewa, except, exie, ezekiel, ezequiel, ezra, f, fabian, fabiola, fae, fairy, faith, fallon, fannie, fanny, far, farah, farrah, fatima, fatimah, faustina, faustino, fausto, fawn, fay, faye, fe, felecia, felica, felice, felicia, felicidad, felicitia, felicitas, felipa, felipe, felisa, felisha, felix, felton, ferdinand, fermin, fermina, fern, fernanda, fernande, fernando, ferne, few, fidel, fidela, fidelia, fifty, filiberto, filomena, fiona, first, five, flavia, fleta, fletcher, flo, flor, flora, florance, florence, florencia, florencio, florene, dottie, florentina, florentino, floretta, floria, florida, florinda, florine, drema, florrie, flossie, drusilla, floy, floyd, dulce, fonda, duncan, for, forest, former, formerly, dwain, forrest, forty, foster, dyan, found, four, fran, france, francene, frances, francesca, francesco, franchisesca, francie, francina, francine, francis, francisca, francisco, francoise, frank, eda, frankie, franklyn, fransisca, fred, freda, edgar, fredda, freddie, edison, freddy, frederic, frederica, frederick, fredericka, fredia, fredric, fredrick, fredricka, freedra, freeman, edwina, freida, frida, frieda, fritz, from, fumiko, eileen, further, g, gabriel, gabriela, gabriele, gabriella, gabrielle, gail, gala, gale, elbert, galen, galina, garfield, garland, garnet, garnett, garret, garrett, garry, gary, gaston, gavin, gay, gaye, elenore, gayla, gayle, gaylene, gaylord, gaynell, gearldine, gema, gemma, gena, gene, genesis, geneva, genevie, genevieve, elin, genevive, genia, genie, genna, gennie, genny, genoveva, geoffrey, georgann, george, georgeann, elizabet, georgeanna, georgene, georgetta, georgette, georgia, georgiana, georgiann, georgianne, georgina, georgine, gerald, geraldine, geraldo, geraldyn, gerard, gerardo, gerda, geri, german, gerri, gerry, gertie, gertrude, gertrudis, get, ghislaine, gia, gianna, gadget, gigi, gilbert, gilberte, gilberto, gilda, gillian, gilma, gina, ginette, ginger, elvin, ginny, gino, giovanna, giovanni, gisela, gisele, giselle, gita, giuseppe, giuseppina, gladis, glady, gladys, glayds, glen, glenda, glendora, glenn, glenna, glennie, glennis, glinda, gloria, glory, glynda, glynis, go, golda, golden, emmett, goldie, gonzalo, good, gordon, got, grace, gracia, gracie, eneida, graciela, grady, graham, grant, granville, grayce, grazyna, great, gregg, gregoria, gregorio, gregory, greta, gretchen, gretta, gricelda, grisel, griselda, guadalupe, gudrun, guillermine, guillermo, gus, gussie, gustavo, guy, gwen, gwenda, gwendolyn, gwenn, gwyn, gwyneth, h, ha, hae, hai, esta, hailey, hal, haley, halina, halley, hallie, han, hang, hanh, hank, hanna, hannah, hannelore, hans, harlan, harland, harley, harmony, harold, harriet, harriett, harriette, harris, harrison, harry, harvey, has, hasnt, hassan, hassie, hattie, have, havent, haydee, eva, hayden, hayley, haywood, hazel, he, heath, heather, hector, hed, hedwig, hedy, heide, heidi, heidy, heike, helaine, helen, helena, helene, helga, hellen, hence, henrietta, henriette, henry, her, herb, herbert, here, hereafter, hereby, herein, heres, hereupon, herlinda, herma, herman, hermelinda, hermina, hermine, faviola, herminia, hers, herschel, herself, federico, hershel, hertha, hes, hester, hettie, hid, hien, hilaria, hilario, hilary, hilda, hildegard, hildred, hillary, hilma, hilton, him, himself, hipolito, hiroko, his, hoa, hobert, holley, holli, hollie, holly, home, homer, honey, hong, hope, horace, horacio, hortencia, hortense, hortensia, hosea, houston, how, howard, however, hoyt, hubert, huey, hugh, hui, hulda, humberto, hundred, hung, hunter, huong, hwa, hye, hyman, hyo, hyon, i, ian, id, ida, idalia, idell, idella, ie, iesha, if, ignacio, franklin, ike, ilana, ileana, ileen, ilene, iliana, ill, illa, ilona, ilse, iluminada, im, ima, imelda, imogene, in, inc, include, includes, indeed, index, india, indira, inell, ines, inez, information, inga, inge, ingeborg, inger, ingrid, inocencia, instead, internet, into, iola, iona, garth, ione, iraida, irena, irene, iris, irish, irma, irmgard, irvin, irving, gaynelle, is, isa, isaac, isabel, genaro, isabell, isabella, isadora, isaiah, isaias, isaura, isela, isiah, isidra, isidro, isis, ismael, isnt, isobel, israel, isreal, issac, it, its, itself, ivan, ivana, ive, georgianna, ivelisse, georgie, ivey, ivonne, ivory, ivy, izetta, izola, j, ja, jacalyn, jacelyn, germaine, jacinda, jacinta, jacinto, gertha, jackeline, gertrud, jackelyn, jacki, gertude, jackie, jacklyn, jackqueline, jackson, jaclyn, gil, jacob, jacqualine, jacque, jacquelin, jacqueline, jacquelyne, jacquelynn, jacques, jacquetta, jacqui, jacquie, jacquiline, jacqueline, jacquelyn, jada, jadowiga, jae, jaime, jaimie, jaimie, jake,

jaleesa, jalisa, jama, jamaal, jamal, jamar, jame, jamee, jamel, james, jamey, jami, jamie, jamika, jamila, jamison, jammie, jan, jana, janae, janay, jane, janean, janee, janeen, graig, janel, janell, janella, janelle, greg, janene, janessa, Janet, janeth, janett, janetta, janette, janey, jani, janice, grover, janie, janina, janine, janis, janise, janita, jann, janna, jannet, jannette, jannie, january, janyce, jaqueline, jaquelyn, jared, jarod, jarred, jarrett, jarrod, jarvis, jasmin, jason, jasper, hana, jaunita, javier, jay, jaye, jayne, jaymie, jayna, jayne, jayson, jazmin, jazmine, jc, jean, jeana, jeane, jeanelle, jeanene, jeanett, jeanetta, jeanette, jeanice, jeanie, jeanine, jeanmarie, jeanna, jeanne, jeannetta, jeannette, jeannie, jeannine, jed, jeff, hee, jefferey, jefferson, jeffery, jeffie, jeffrey, jeffry, jen, jena, jenae, jene, jenee, jenell, jenelle, jenette, jeneva, heriberto, jeni, jenice, jenifer, jeniffer, hermila, jenine, jenise, jenna, jennefer, jennell, herta, jennette, jenni, jennie, hiedi, jennifer, jenniffer, jennine, jerald, jeraldine, hilde, jeramy, hildegard, jere, jeremiah, jeri, jericia, jerilyn, hiram, jermaine, hisako, jerold, jerome, jeromy, jerrell, jerri, hollis, jerrica, jerrie, jerrod, jerrold, jerry, jesica, jess, jesse, jessenia, jessi, jessia, jessica, jessika, jestine, hsiu, jesus, hue, jesusa, jesuita, hugo, jetta, jettie, jewel, jewell, ji, jill, jillian, hyacinth, jim, jimmie, jin, jina, hyun, jinny, jo, joan, joana, joane, joann, ignacia, joanna, joanne, ila, joannie, ilda, joaquin, joaquina, jocelyn, jodee, jodi, jodie, jody, joeann, joel, joella, joelle, joellen, ina, joesph, joetta, joette, joey, johana, johanna, johanne, john, johna, johnathan, johnathon, johnetta, johnie, johnna, ira, johnnie, johnny, johnsie, irina, johnson, joi, joie, jolanda, joleen, jolene, irwin, jolie, joline, jolyn, jolynn, jon, isabelle, jona, jonah, jonas, jonathon, jone, jonell, jonelle, jong, joni, jonie, jonna, jonnie, jordan, jordon, iva, jorge, jose, josef, ivette, josefa, josefina, josefine, joselyn, joseph, josephina, josephine, josette, josh, joshua, josiah, josie, jack, joslyn, jospeh, josphine, josue, jovan, joy, joya, joyce, joycelyn, joye, juan, juana, juanita, jacquelyn, jude, judi, judie, judith, judson, judy, julee, julene, jules, juli, jade, julia, julian, juliana, juliane, juliann, julianna, julianne, julie, julieann, julienne, juliet, julieta, julietta, juliette, julio, julissa, julius, june, jung, junie, junior, junita, junko, just, justa, justina, justine, jutta, k, ka, kacey, kaci, kacie, kacy, kai, kaila, kaitlin, kaitlyn, kala, kaleigh, kaley, kali, kallie, kalyn, kam, kamala, janiece, kami, kamilah, kandace, kandice, kandis, kandra, kandy, kanesha, kanisha, kara, karan, kareem, kareen, karen, karenia, karey, kari, karie, karima, karin, karina, jasmine, karine, karisa, karissa, karl, karla, karleen, karlene, karly, karlyn, karma, karmen, karol, karole, karoline, karolyn, karon, karren, karri, karrie, karry, kary, karyl, karyn, kasandra, kasha, kasi, kassandra, kassie, kate, katelin, katelyn, katelynn, katherine, kathaleen, katharina, katharine, katharyn, kathe, katheleen, katherin, katherine, kathern, kathey, kathi, kathie, kathleen, kathlene, kathline, kathlyn, kathrin, kathrine, kathryn, kathryne, kathy, kathryn, kati, katia, katie, katlyn, katrice, katrina, kattie, katy, kay, jenny, kayce, kaycee, kaye, kayla, kaylee, jeremy, kayleen, kayleigh, kaylene, jerylene, kazuko, kecia, keeley, keely, keena, keesha, keiko, keila, keira, keisha, keith, jesenia, keitha, keli, kelle, kelley, kelli, kellie, kelly, jessie, kellye, kelsey, kelsi, kelsie, kelvin, kemberly, ken, kena, kenda, kendal, kendall, kendra, kendrick, keneth, jimmy, kenia, kenisha, kenna, kenneth, kenny, kent, kenton, joanie, kenyatta, kenyetta, kera, keren, keri, kermit, kerri, kerrie, kerry, kerstin, kesha, joe, keshia, keturah, keva, keven, kevin, khadijah, khalilah, kia, kiana, kiara, kiera, kiersten, kiesha, kieth, kim, kimber, kimberlee, johnette, kimberley, kimberlie, kimberly, kimberly, kimbra, kimi, kimiko, kina, kindra, king, kira, kirby, kirk, kirsten, kirstie, kirstin, kisha, kit, kittie, jonathan, kitty, kiyoko, kizzie, kizzy, klara, know, korey, kori, kortney, kory, kourtney, kraig, kris, krishna, krissy, krista, kristal, kristan, kristeen, kristel, kristen, kristi, kristian, kristie, kristin, kristina, kristine, kristofer, kristy, kristyn, krysta, jovita, krystal, krysten, krystin, krystina, krystle, krystyna, kum, kurt, kurtis, kyla, kyle, kylee, kylie, kym, jule, kymberly, kyoko, kyong, kyra, kyung, l, lacey, lachelle, laci, lacie, lacresha, lacy, ladawn, ladonna, lady, lael, lahoma, lai, laine, lajuana, lakeesha, lakeisha, lakendra, lakenya, lakesha, lakeshia, lakia, lakiesha, justin, lakisha, lakita, lala, lamonica, lamont, lan, lana, lance, landon, lane, lanell, lanelle, lanette, lani, lanie,

lanita, lannie, lanny, lanora, laquanda, laquita, lara, larae, kandi, laraine, laree, larhonda, larisa, larissa, larita, laronda, larraine, larry, larue, lasandra, lashanda, lashandra, lashaun, lashaunda, lashawn, lashawna, lashay, lashell, lashon, lashonda, lashunda, last, latanya, latasha, latashia, later, latesha, latia, laticia, latina, latisha, latonia, latonya, latoria, latosha, latoya, latoyia, latrice, latricia, latrina, latrishia, kasey, latter, latterly, kasie, launa, laura, lauralee, lauran, laure, laureen, laurel, Lauren, Laurena, Laurence, Laurene, Lauretta, Laurette, Lauri, Katherina, Laurice, Laurie, Katheryn, Laurinda, Laurine, Lauryn, Lavada, Lavelle, Lavenia, Lavera, Lavern, Laverna, Laverne, Laveta, Lavette, Lavinia, Lavon, Lavona, Lavonda, Katina, Lavone, Lavonia, Lavonna, Lawana, Lawanda, Lawanna, Lawrence, Lawrence, Layla, Layne, Lazaro, Le, Lea, Leah, Lean, Leana, Leandra, Leandro, Leann, Keenan, Leanna, Leanne, Leanora, Least, Leatha, Leatrice, Lecia, Leda, Leeann, Kellee, Leeanna, Leeanne, Leena, Leesa, Left, Leia, Leida, Leif, Leigh, Leigha, Leighann, Leila, Leilani, Leisa, Leisha, Lekisha, Lela, Lelah, Leland, Lelia, Lemuel, Len, Kenneth, Lena, Lenard, Lenita, Kenya, Lenna, Lennie, Lenny, Lenora, Lenore, Leo, Leola, Leoma, Leon, Leona, Leonard, Leonarda, Leonardo, Leone, Lionel, Leonia, Leonida, Leonie, Leonila, Leonor, Leonora, Leonore, Leontine, Leopoldo, Leora, Kiley, Leota, Lera, Kimberely, Leroy, Les, Kimberli, Lesa, Lesha, Lesia, Leslee, Lesley, Leslie, Less, Lessie, Kip, Lester, Let, Leta, Letha, Leticia, Letisha, Letitia, Lets, Lettie, Letty, Levi, Lewis, Lezlie, Li, Lia, Liana, Liane, Lianne, Libbie, Libby, Liberty, Librada, Lida, Lidia, Lien, Lieselotte, Ligia, Like, Likely, Lila, Lily, Lilia, Lilian, Liliana, Lilla, Kristle, Lilli, Kristopher, Lillia, Lillian, Lillian, Lillian, Lillie, Lilly, Lily, Lin, Lina, Lincoln, Linda, Lindsay, Lindsey, Lindsay, Lindy, Line, Linette, Ling, Linh, Links, Linn, Linnea, Linnie, Lino, Linsey, Linwood, Lionel, Lisa, Lisabeth, Lisandra, Lisbeth, Lise, Lisette, Lisha, Laila, Lissa, Lisette, Lita, Livia, Liz, Liza, Elizabeth, Lizbeth, Lizeth, Lizette, Lizzette, Lizzie, Il, Lamar, Lloyd, Loan, Logan, Loida, Lois, Lola, Lolita, Loma, Lon, Lona, Lang, Londa, Loni, Lonna, Lonnie, Lonny, Lora, Loraine, Lorealee, Lore, Lorean, Loree, Loreen, Lorelei, Loren, Lorena, Lorene, Lorenza, Lorenzo, Loreta, Loretta, Lorette, Lori, Loria, Loriann, Lorie, Lorilee, Lashawnda, Lorinda, Loris, Lorita, Lorna, Lorraine, Lasonya, Lorretta, Latarsha, Lorri, Lorriane, Lorrie, Lorrine, Lory, Lottie, Lou, Louanne, Louella, Louetta, Louie, Louis, Louisa, Louise, Loura, Lourdes, Lourie, Love, Lovella, Lovetta, Lovie, Lowell, Loyce, Loyd, Ltd, Lu, Luana, Luann, Luanna, Luanne, Luba, Lucas, Luci, Lucia, Luciano, Lucie, Lucien, Lucienne, Lucila, Lucilla, Lucille, Lucina, Lucio, Lucius, Lucrecia, Lavina, Lucretia, Lucy, Ludie, Ludivina, Luella, Luetta, Luigi, Lavonne, Luis, Luisa, Luise, Luke, Lula, Lulu, Luna, Lupe, Lupita, Lura, Lurlene, Lurline, Luther, Luvenia, Luz, Lyda, Lydia, Lyla, Lyle, Lyman, Lyn, Lynda, Lyndia, Lee, Lyndon, Lyndsai, Lyndsey, Lynell, Lynelle, Lynette, Lynn, Lynna, Lynne, Lynnette, Lynsey, Lynwood, Mabel, Mabelle, Mable, Mac, Machel, Macie, Mack, Mackenzie, Macy, Madalene, Madaline, Madalyn, Maddie, Made, Madeline, Madeleine, Madelene, Madeline, Madelyn, Madge, Madie, Madison, Madlyn, Madonna, Mae, Maegan, Mafalda, Magali, Magaly, Magan, Margaret, Magda, Magdalen, Magdalena, Magdalene, Magen, Maggie, Magnolia, Mahalia, Mai, Maia, Maida, Maile, Maira, Maire, Maisha, Maisie, Major, Majorie, Make, Makeda, Makes, Malcolm, Malcom, Malena, Malia, Malik, Malika, Malinda, Malisa, Lexie, Malissa, Malka, Mallie, Mallory, Malorie, Mamie, Mammie, Man, Mana, Manda, Mandi, Mandie, Mandy, Manie, Manual, Manuel, Manuela, Many, Maple, Mara, Maragaret, Maragret, Maranda, Marc, Marcel, Marcela, Marcelina, Marceline, Marcelino, Marcell, Marcella, Marcelle, Marcellus, Marcelo, Marcene, Marchelle, Marcia, Marcie, Marcos, Marcus, Marcy, Maren, Marg, Margareta, Margaret, Margaretta, Margarette, Margarita, Margarite, Margarito, Margart, Margene, Margeret, Margert, Margery, Marget, Margherita, Margie, Margit, Margo, Margorie, Margret, Margrett, Marguerita, Marguerite, Margurite, Margy, Martha, Mari, Maria, Loise, Mariah, Mariam, Marian, Mariana, Marianela, Mariann, Marianna, Marianne, Mariano, Maribel, Maribeth, Marica, Maricela, Maricruz, Marie, Mariel, Mariela, Mariella, Marietta, Mariette, Mariko, Marilee, Marilou, Marilu, Marilyn, Marilynn, Marin, Marina, Marinda, Marine, Mario, Marion, Lorina, Maris, Lorine, Marisa, Marisela, Marisha, Marisol, Marissa, Marita, Maritza, Marivel, Marjory, Mark, Marketta, Markita, Louann, Markus, Marla, Marlana, Marleen, Marlen, Marlena, Marlene, Marlin,

marline, marlo, louvenia, marlon, marlyn, marlys, marna, marni, marnie, marquerite, marquette, marquis, marquita, marquitta, marry, marsha, marshall, marth, martha, luciana, marti, martin, martina, martine, marty, lucile, marva, marvel, marvella, lucinda, marvin, marvis, marx, mary, marya, maryalice, maryam, lue, maryann, maryanna, maryanne, marybelle, marybeth, maryellen, maryetta, maryjane, maryjo, maryland, marylee, marylin, maryln, marylou, marylouise, marylyn, marylynn, maryrose, masako, mason, matha, mathew, mathilda, mathilde, matilda, matilde, matt, matthew, mattie, maud, maude, lynetta, maudie, maura, maureen, maurice, mauricio, maurine, maurita, ma, mauro, mavis, max, maxie, maxima, maximina, maximo, maxine, maxwell, maya, maybe, maybell, maybelle, maye, mayme, maynard, mayola, mayra, mazie, mckenzie, mckinley, me, meagan, meaghan, meantime, meanwhile, mechelle, meda, mee, meg, megan, meggan, megghan, meghann, mei, mel, melaine, melani, melania, melanie, melany, melba, melda, melia, melida, melina, melinda, melisa, melissa, melissia, melita, mellie, mellisa, mellissa, melodee, melodi, melodie, melody, melonie, melony, melva, malvina, melvin, melvina, mendy, mercedes, mercedez, mercy, meredith, meri, merideth, meridith, merilyn, merissa, merle, mao, merlene, merlin, merlyn, merna, merri, merrie, merrilee, merrill, marcelene, merry, mertie, mervin, meryl, meta, mi, mia, mica, micaela, micah, marci, michael, michaela, marco, michaele, michal, michale, mardell, micheal, michel, margaret, michele, margarete, michelina, micheline, michell, michelle, michiko, mickey, micki, marge, mickie, miesha, migdalia, mignon, miguelina, mika, mikaela, mike, mikel, miki, margot, mikki, milagro, milagros, milan, milda, mildred, miles, milford, milissa, millard, millie, million, milly, milo, milton, mimi, min, mina, minda, mindi, mindy, minerva, ming, minh, minna, minnie, minta, marielle, miquel, mira, miranda, mireille, mireya, miriam, mirian, mirna, mirta, mirtha, misha, miss, missy, misti, mistie, misty, mitch, mitchel, mitchell, mitsue, mitsuko, mittie, marjorie, mitzi, mitzie, miyoko, modesto, mohamed, mohammad, mohammed, moira, mollie, molly, moment, mona, monet, monica, monika, monique, monnie, monroe, monserrate, monte, monty, moon, mora, more, moreover, morgan, moriah, morris, marta, morton, mose, moises, moshe, most, mostly, mozell, mozella, mozelle, mr, mrs, much, mui, muoi, murray, must, my, myesha, myles, myong, myra, myrl, myrle, myrna, myron, myrta, myrtie, myrtle, myself, myung, n, nada, nadene, nadia, naida, nakesha, nakia, nakisha, nakita, nam, namely, nan, nana, nancee, nancey, nanci, nancie, nancy, nanette, nannette, nannie, naoma, naomi, napoleon, narcisa, natacha, natalia, natalie, natalya, natasha, natashia, nathalie, nathanael, nathanial, natisha, natividad, natosha, neal, near, necole, ned, neda, nedra, neely, neida, neil, neither, nelda, nelia, nelida, nell, nella, nelle, nellie, nelly, nelson, nena, nenita, neoma, neomi, nereida, nerissa, nery, nestor, neta, nettie, neva, nevada, never, nevertheless, neville, new, newton, next, nga, ngan, ngoc, nguyet, nia, nichelle, nichol, nicholas, nichole, melynda, nicholle, nick, nicki, nickie, nickole, nicky, nicol, nicola, nicolas, nicolasa, nicole, nicolette, nicolle, nida, nidia, niesha, nieves, nigel, niki, nikia, nikita, nikki, nikole, nila, nilda, nilsa, nina, nine, ninety, micha, ninfa, nisha, nita, no, noah, noble, nobody, nobuko, noe, noel, noelia, noella, noelle, noemi, nola, nolan, noma, nona, none, miguel, nonetheless, noone, nor, nora, norah, norbert, norberto, mila, noreen, noriko, norine, norma, norman, normand, not, nothing, nova, millicent, novella, now, nowhere, nu, nubia, numbers, nydia, nyla, o, obdulia, ocie, octavia, octavio, oda, odelia, odell, odessa, odette, odilia, odis, mirella, of, ofelia, off, often, oh, ola, olen, olene, oleta, olevia, olga, olimpia, olin, olinda, oliva, olive, oliver, olivia, ollie, olympia, oma, modesta, omar, omega, omer, omitted, on, moises, ona, once, one, oneida, ones, onie, onita, only, onto, opal, ophelia, or, oralee, oralia, ord, oren, oretha, orlando, orpha, orval, orville, oscar, ossie, osvaldo, oswaldo, otelia, muriel, otha, other, others, otherwise, otilia, otis, myriam, otto, ouida, our, ours, ourselves, myrtice, out, myrtis, over, owen, own, ozell, ozella, ozie, nadine, p, pa, pablo, page, pages, paige, palma, palmer, palmira, pamala, pamela, pamelia, pamella, pamila, pamula, pansy, paola, paris, parker, part, parthenia, particia, pasquale, pasty, pat, patience, nathan, patria, patrica, nathaniel,

sherilyn, sherise, sherita, sherlene, sherley, sherly, sherlyn, sherman, sheron, sherrell, sherri, sherrie, sherril, savannah, sherrill, sherron, sherry, sherryl, sherwood, shery, sheryl, sheryll, shes, shiela, shila, shiloh, shin, shira, shirely, shirl, shirlee, shirleen, shirlene, shirley, shirly, shizue, shizuko, shon, shona, setsuko, shonda, shondra, shonna, shonta, shoshana, shu, shakira, shyla, sibyl, sid, sidney, sierra, signe, sigrid, silas, silva, silvana, silvia, sima, simon, simona, simone, simonne, sina, since, shani, sindy, siobhan, sirena, siu, shanna, six, sixty, skye, sylvia, so, socorro, sofia, soila, sol, solange, soledad, solomon, some, somehow, someone, somer, something, sometime, sometimes, somewhere, sommer, son, sona, sharita, sondra, song, sonia, sonja, sonny, sonya, sharonda, soo, sharron, sook, soon, sophia, sophie, soraya, sparkle, spencer, spring, stacee, stacey, staci, stacia, stacie, stacy, stanford, shawna, stanley, stanton, star, starla, starr, stefan, stefani, stefania, stefanie, stefany, steffanie, stella, stepanie, stephan, stephane, stephani, stephania, stephanie, stephen, stephenie, stephine, stephnie, sterling, shelton, steve, steven, stevie, stewart, still, stop, stormy, stuart, suanne, such, sudie, sue, sueann, suellen, suk, sulema, sumiko, summer, sun, sunday, sung, sunni, sunny, sunshine, susan, susana, susann, susanna, susannah, susanne, susie, susy, suzan, suzann, suzanna, susanne, suzette, suzi, suzy, svetlana, sybil, syble, sydney, sylvester, sylvia, sylvie, synthia, syreeta, t, ta, tabetha, tabitha, tad, tai, taina, taisha, tajuana, takako, taking, takisha, talia, talisha, talitha, tama, tamar, tamara, tamatha, tambra, tameika, tameka, tamekia, tamela, tamera, tamesha, tami, tamica, tamie, tamiko, tamisha, sixta, tammara, tammara, tammi, tammie, tammy, tana, tandra, tandy, taneka, tanesha, tangela, tania, tanika, tanja, tanna, tanner, tara, tarah, taren, tarra, tarsha, taryn, tasha, tashia, tashina, tasia, tatiana, tatum, tatyana, taunya, tawana, tawanda, tawanna, tawna, tawnya, stan, taylor, tayna, ted, teddy, tegan, tell, stasia, telma, temeka, temika, tempie, temple, ten, tena, tenesha, stephaine, tenisha, tennie, tennille, teodora, teodoro, stephany, teofila, tequila, tera, tereasa, terence, teresa, teresia, teresita, teressa, teri, terica, su, terina, terisa, terra, terrance, terrell, terrence, terresa, terri, terrie, terrilyn, terry, tesha, tessa, tessie, thad, thaddeus, thalia, than, thanh, thao, that, thatll, thats, thatve, the, theda, their, thelma, them, suzie, themselves, then, thence, theo, theodora, theodore, theola, thereafter, thereby, thered, therefore, tabatha, therein, therell, therere, theres, theresa, therese, theressa, thereupon, thereve, thearon, thersa, these, tam, theyd, tamala, theyll, theyre, theyve, thi, thing, think, thirty, this, thomas, thomasena, thomasina, thomasine, thora, tamika, those, though, thought, thousand, thresa, through, throughout, tamra, thru, thu, thurman, thus, thuy, tia, tiana, tianna, tanisha, tiara, tien, tieria, tanya, tierra, tiesha, tifany, tari, tiffaney, tiffani, tiffanie, tiffany, tiffiny, tijuana, til, tilda, till, tillie, tim, timika, timmy, timothy, tina, tawny, tinisha, tiny, tip, tisa, tish, teena, tisha, teisha, titus, to, tobi, tobias, tobie, toby, toccara, tod, together, toi, told, tom, tomas, tomasa, tomeika, tomi, tomika, tomiko, terese, tommie, tommy, tommye, tomoko, tona, tonda, tonette, toney, tonia, tonie, tonisha, tonita, tonja, tony, tonya, too, tora, tess, tori, torie, torri, torrie, tory, tosha, toshia, thea, toshiko, tova, towanda, towards, toya, tracee, traci, tracie, theresia, tracy, tran, trang, travis, treasa, treena, trena, trent, trenton, tresa, tressa, tressie, treva, trevor, trey, tricia, trillion, trina, trinh, trinidad, trinity, trisha, trista, tristan, troy, trudi, trudie, trudy, trula, truman, try, tu, tuan, tula, tuyet, twana, twanda, twanna, twenty, twila, two, twyla, ty, tyesha, tyisha, tyler, tynisha, todd, tyra, tyree, tyron, tyrona, tyson, u, ulrike, ulysses, un, una, under, unless, unlike, until, unto, up, toni, upon, ursula, us, used, usha, using, ute, v, vada, val, valarie, valda, valencia, valene, valentin, valentina, valentine, valeri, valeria, valerie, tracey, valery, valorie, valrie, van, vanda, vanesa, vanessa, vanetta, vania, vanita, vanna, vannesa, vannessa, vashti, vasiliki, vaughn, ve, veda, velda, velia, vella, velma, trish, velva, velvet, vena, venessa, venetta, venice, venita, vennie, venus, veola, vera, verda, verdell, verdie, verena, vergie, verla, verlene, verlie, verline, vern, verna, vernell, vernetta, vernia, tyrell, vernice, vernie, vernita, ula, vernon, verona, veronica, veronika, veronique, versie, vertie, very, vesta, veta, vi, via, vicenta, vicente, vickey, vickie, vicky, victor, victoria, vida, vallie, viki, vikki, vilma,

vance, vina, vince, vincent, vincenza, vincenzo, vinita, vinnie, viola, violet, violeta, violette, virgen, virgie, virgil, virgilio, virgina, virginia, vita, vito, viva, vivan, vivian, viviana, vivien, vivienne, vol, vols, von, voncile, vonda, vonnie, vs, w, wade, wai, waldo, walker, wallace, wally, walter, walton, waltraud, wan, wanda, waneta, wanetta, wanita, ward, warner, warren, was, wasnt, wava, way, waylon, wayne, we, wed, vicki, wei, weldon, well, wen, victorina, wendell, wendi, wendie, wendolyn, wendy, wenona, went, were, werent, werner, wes, wesley, weston, weve, what, whatever, whatll, whats, whatve, when, whence, whenever, where, whereafter, whereas, whereby, wherein, wheres, whereupon, wherever, whether, which, while, whim, whither, whitley, whitney, who, whod, whoever, whole, wholl, whom, whomever, whos, whose, why, wilber, wilbert, wilbur, wilburn, wilda, wiley, wilford, wilfred, wilfredo, wilhelmina, wilhemina, will, willa, willard, willena, willene, willetta, willette, willia, william, williams, willian, willie, williemae, willis, willodean, willow, willy, wilma, wilmer, wilson, wilton, windy, winford, winfred, winifred, winnie, winnifred, winona, winston, winter, with, within, without, wm, wonda, wont, woodrow, words, wouldnt, wyatt, wynell, wynona, x, xavier, xenia, xiao, xiomara, xochitl, xuan, y, yadira, yaeko, yael, yahaira, yajaira, yan, yang, yanira, yasmin, yasmine, yasuko, yee, yelena, yen, yer, yes, yesenia, yessenia, yet, yetta, yevette, yi, ying, yoko, yolanda, yolande, yolando, yolonda, yon, yong, yoshie, yoshiko, you, youd, youlanda, youll, young, your, youre, yours, yourself, yourselves, youve, yu, yuette, yuk, yuki, yukiko, yuko, yulanda, yun, yung, yuonne, yuri, yuriko, yvette, yvone, yvonne, z, zachariah, zachary, zachery, zack, zackary, zada, zaida, zana, zandra, zane, zelda, zella, zelma, zena, zenaida, zenia, zenobia, zetta, zina, zita, zoe, zofia, zoila, zola, zona, zonia, zora, zoraida, zula, zulema, zulma