- i162 Analyser et modéliser des données
 - Données ≠ informations
 - Analyser et modéliser des données
- Stocks de données
 - Stocks de données
 - Structurés ou non structurés
- Code et codage
 - Code ≠ codage
- Format de données
 - Qu'est-ce qu'un format pour une donnée en informatique ?
 - Différents formats de données
 - o Format de numéro téléphone
 - Format File signature :
 - <u>Table ASCII (American Standard Code for Information Interchange)</u>

• Conversion

- Programme en Python généré par "chat GPT"
 pour faire un fichier de dates (iso 8601) to
 (format csv):
- Changer le titre d'un album, le nom de l'artiste dans un fichier .mp3 à l'aide d'un éditeur hexadécimal.
- EXCEL format des dates
- Format d'un mail (RFC 5321)
- <u>Format d'une url (RFC 3986)</u>
 <u>https://fr.wikipedia.org/wiki/Uniform_Resource_</u>
 <u>e Locator</u>
- Qu'est ce que la collation du point de vue des caractères en informatique
- DNS

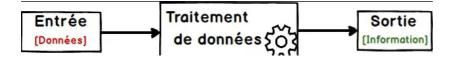
- O Qubit vs Bit
- Quantique
- Entité

i162 - Analyser et modéliser des données

Loré Godel 06/09/23

Données ≠ informations

Les données ne sont PAS des informations!



Définition:

- Données : contiennent des chiffres, des énoncés et des caractères sous forme brute.
- **Information** : une nouvelle, un renseignement, une documentation sur quelque chose ou sur quelqu'un, portés à la connaissance de quelqu'un.

Exemples:

- Données :
 - Prénom
 - Nom
 - Age
 - Nationalité
 - Médicales
 - Scientifiques
 - Réseaux sociaux (likes, commentaires, partages)
- Information : Données traitées, organisées, structurées ou présentées dans un contexte donné afin de les rendre utiles -> "Il fait chaud aujourd'hui!" source image :

(<u>https://waytolearnx.com/2018/08/difference-entredonnee-et-information.html</u> (https://waytolearnx.com/2018/08/difference-entredonnee-et-information.html

donnee-et-information.html)

sources text:

Données (https://fr.wikipedia.org/wiki/Donnée (informatique)))

Analyser et modéliser des données



Défnition:

- **Analyser des données** : Examiner/Interpréter les données accompagné de graphiques pour élaborer des réponses à des questions.
- Modéliser des données : Etablire la structure, les associations, les relations et les contraintes des données disponibles.

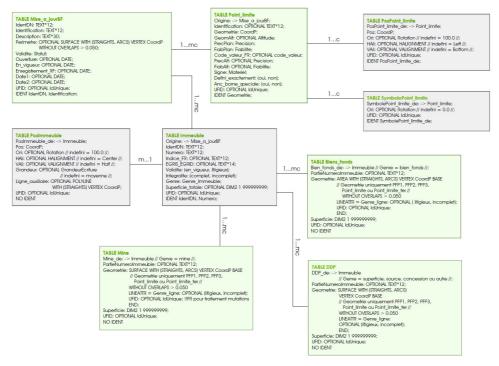
Exemples:

• Analyse de base de données ex.:

Tableau de bord du rapport d'analyse des accidents de NYC This side shows the purples of vasines medicin which helps in analyzing the clash of accidents report of New York City. Data covered under this decidence of the accident kend over time, number of persones injurised by the state of accidents with Killed Accident

This graph/chart is linked to excel, and changes automatically based on data. Just left click on it and select "edit data"

• Modélisation base de données ex.:



sources images:

 $(\underline{https://www.imope.fr/data.html}_{(https://www.imope.fr/data.html)}) \\ (\underline{https://www.slideteam.net/blog/modeles-de-rapport-danalyse-ppt?lang=French}_{(https://www.slideteam.net/blog/modeles-de-rapport-danalyse-ppt?lang}) \\ \\ (\underline{https://www.slideteam.net/blog/modeles-de-rapport-danalyse-ppt?lang}) \\ \\ (\underline{https://www.slideteam.net/blog/modeles-de-rapport-danalyse-ppt?lang}) \\ (\underline{https://www.slideteam.net/blog/modeles-de-rapport-danalyse-ppt.d$

danalyse-ppt?lang=French)

(https://www.fr.ch/territoire-amenagement-et-geomatique/modeles-de-donnees (https://www.fr.ch/territoire-amenagement-et-geomatique/modeles-de-donnees

et-constructions/cartes-plans-cadastre-et-geomatique/modeles-de-donnees)

sources text:

(https://www.talend.com/fr/resources/guide-modelisation-donnees/#:~:text=La modélisation des

données (https://www.talend.com/fr/resources/guide-modelisation-

donnees/#:~:text=La%20mod%C3%A9lisation%20des%20donn%C3%A9es%20)

Stocks de données

Stocks de données

Evitons la redondance!

- **Pourquoi? :** Pour réduit le gaspillage d'espace de stockage, améliorer la maintenance et assurer la cohérence des données.
- Exemple: au lieu de stocker plusieurs fois la même information, une base de données peut utiliser des clés uniques pour référencer des données partagées, ce qui économise de l'espace et facilite les mises à jour sans avoir à les répéter.



CD, HDD, SSD, USB...!



Ex.: Dropbox, Google Drive, et Amazon S3



Data center

Exemples:

- Stockage de médias (musique, films et photos numériques)
- Stockage de fichiers personnels (enregistrer un document word sur son pc pour qu'il reste le même jusqu'à la prochaine modification)
- Données de télémétrie (Les entreprises technologiques collectent des données de télémétrie à partir de leurs produits et services, telles que l'utilisation des applications, les performances des appareils, etc., pour améliorer leurs produits.)

```
# Stock de données

## précis

### DEF: Grande attention aux détails et à la précision

### EX: Carte de crédit.

### DEF: Large éventail de données et de détails sur un sujet particulier.

### EX:Données démographiques d'une population

## redondant

### DEF: Données répétées inutilement, créant potentiellement des incohérences

### Ex: Répétition inutile des données d'un client

## incohérent

### DEF: Données contradictoires ou non conformes aux normes, entraînant une imprécision

### EX: Dates de naissance au format différent (ex. MM-JJ-AAAA et AAAA-MM-JJ) dans la ba
```

Exemple visu.:

exhaustif:

ID Produit	Nom du Produit	Catégorie	Marque	Poids (g)	Prix (EUR)	Date d'Expirati
1	Pommes	Fruits	Ferme X	150	1.99	2023- 09-30
2	Lait	Produits laitiers	Marque A	1000	0.99	2023- 10-15

• redondant:

ID	Nom	Adresse	Email	Téléphor
1	John Doe	123 Rue de la Ville	john@example.com	555-123- 4567
2	Jane Smith	456 Avenue du Quartier	jane@example.com (mailto:jane@example.com)	555-987- 6543

Les données du client (Nom, Adresse, Email, Téléphone) sont stockées à la fois dans cette table et dans la table "Commandes", créant une duplication inutile.

CommandelD	ClientID	Date de Commande	Total
101	1	2023-09-15	50.00
102	2	2023-09-16	30.00

• incoérent :

ID	Nom	Date de Naissance		
1	John Doe	05-15-1990		
2	Jane Smith	1992-07-20		
3	Alice Lee	12/04/1985		
4	Bob White	1988-10-30		

sources images:

 $(\underline{https://quizizz.com/admin/quiz/5e3b6f663bd6ad001b}\\ \underline{d2679a/storage-media-questions-from-students-cie-}\\ \underline{ict-igcse-unit-3}_{(\underline{https://quizizz.com/admin/quiz/5e3b6f663bd6ad001bd2679a/storage-media-likely-$

questions-from-students-cie-ict-igcse-unit-3)

(<u>https://medium.com/meet-lima/whats-the-cloud-what-s-a-private-cloud-7fcdd8b535d1 (https://medium.com/meet-lima/whats-the-cloud-</u>

lima/whats-the-cloud-what-s-a-private-cloud-7fcdd8b535d1)

(<u>https://home.cern/fr/news/news/computing/cern-data-centre-passes-100-petabytes</u>

(https://home.cern/fr/news/news/computing/cern-data-centre-passes-100-petabytes)

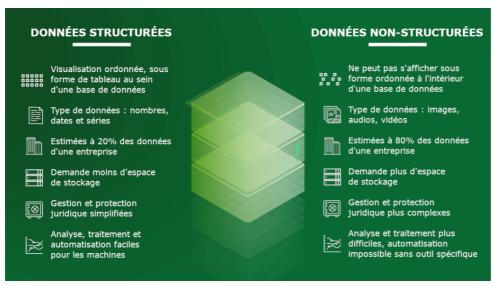
sources texte:

(https://fr.differbetween.com/article/what is the difference between data redundancy and data inconsisten cyd

(https://fr.differbetween.com/article/what is the difference between data redundancy and data inconsistencyd) prompt ChatGPT:

"stock de données exhaustif exemple" "peux-tu me donner un exemple de stock de données redondant ?" "peux-tu me donner un exemple de stock de données incohérent ?" "quel est le meilleur exemple de stock de données exhaustif?"

Structurés ou non structurés



Définition:

- Données structurés: Organisées de manière très formelle et tabulair, elles sont généralement stockées dans des bases de données relationnelles ou des feuilles de calcul.
- Données non structurées: Sans format fixe ou de structure préalablement définie, elles peuvent prendre la forme de texte brut, de médias tels que des images

ou des vidéos.

Exemple:

• structuré:

- carte d'identité
- factures
- tableaux de bord financiers
- bases de données de ressources humaines

• non structuré:

- o signature sur la carte d'identitée
- o Articles de blog
- vidéos YouTube
- o images Instagram
- o enregistrements audio de conversations
- courriels

source image:

(<u>https://www.group-edt.fr/les-differents-types-de-</u>

 $\underline{\textbf{donnees/}}_{(\text{https://www.group-edt.fr/les-differents-types-de-donnees/)}})$

Code et codage

Code ≠ codage

Aspect	Code	Codage
Nature	Instructions	Transformation des informations
Language	Langage de programmation	Format ou représentation
But	Indiquer à l'ordinateur quoi faire	Rendre les données compréhensibles par l'ordinateur

Définition:

 code: mots-clés/ symboles/chiffres et structures logiques qui permettent à toute sorte de dispositifs électroniques de comprendre les tâches à accomplir et de les éxecuter.

- Dispositifs éléctroniques capables de comprendre/d'éxecuter du code:
 - Ordinateurs personnels (PC et Mac)
 - Serveurs
 - Superordinateurs
 - Automates et dispositifs embarqués
 - Calculatrices
 - IoT (Internet des objets)
- codage: Permet de représenter l'information de manière à ce qu'elle puisse être traitée, stockée, transmise et interprétée par des systèmes informatiques.

Les données sont généralement stockées sous forme de séquences de bits (0 et 1) dans les systèmes informatiques.

Exemple:

- Code:
 - Open Python :

```
# Ceci est un commentaire en Python
nom = "Alice"
print("Bonjour, " + nom + "!")
```

affiche "Bonjour, Alice!" à l'écran.

Fortran (Formula Translation) :

```
program bonjour
    print *, "Bonjour, Monde!"
end program bonjour
```

COBOL (Common Business-Oriented Language):

```
IDENTIFICATION DIVISION.

PROGRAM-ID. BonjourMonde.

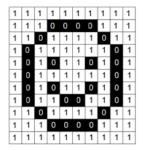
PROCEDURE DIVISION.

DISPLAY 'Bonjour, Monde!'.
```

• HTML:

• Codage:

- L'encodage des caractères textuels en binaire (comme UTF-8 pour les caractères Unicode)
- la compression d'images JPEG
- la conversion d'audio en MP3



Codage d'une image.

Source image:

(<u>https://www.loutrel.fr/wikisn/doku.php?</u> <u>id=les exposes:image format d une image bmp</u>

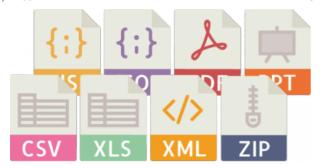
(https://www.loutrel.fr/wikisn/doku.php?id=les_exposes:image_format_d_une_image_bmp)

Format de données

Qu'est-ce qu'un format pour une donnée en informatique ?

L'extension d'un fichier n'est PAS le format d'un fichier

Différence : l'extension de fichier indique le type de fichier, tandis que le format de fichier spécifie la manière dont les données sont organisées et encodées à l'intérieur du fichier.



Définition : Le format de données = format de fichier, est une spécification qui définit la structure, l'encodage et l'organisation des données à l'intérieur du fichier. Il est essentiel pour l'interopérabilité entre les logiciels et les systèmes, ainsi que pour garantir que les données sont lues et interprétées correctement.

Exemple:

- JPEG pour les images
- MP3 pour l'audio
- JSON pour l'échange de données
- CSV pour les données tabulaires

Il existe beaucoup de formats...

source image:

(http://editorialisation.org/ediwiki/index.php/Accueil

(http://editorialisation.org/ediwiki/index.php/Accueil)

Différents formats de données



Chaque format de données est adapté à des besoins particuliers en termes de stockage, de transmission et de traitement. Voici quelques-uns des formats de données les plus courants :

Catégorie	Format	Description
Documents	PDF	Portable Document Format, utilisé pour les documents avec mise en page fixe.
	DOCX	Format de fichier de document Microsoft Word.
	ODT	Open Document Text, format ouvert de traitement de texte.
Base de données	SQL	Structured Query Language, utilisé pour interagir avec les bases de données relationnelles.
	NoSQL	Divers formats de base de données non relationnelles, tels que MongoDB, Cassandra, etc.
	Shapefile	Format de données géographiques vectorielles largement utilisé.

Source image:

(https://www.eazydraw.com/fileFormats.htm

(https://www.eazydraw.com/fileFormats.htm)

Prompt chat gpt:

"C'est quoi exactement et techniquement le format des donées ?""

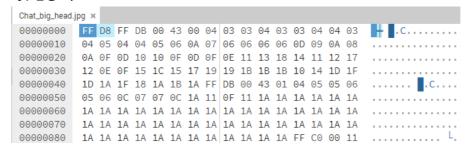
Format de numéro téléphone

Pays/zone géographique	Indicatif de pays	Préfixe international	Préfixe national	Numéro national (significatif)	UTC/DST*	Note
Sao Tomé-et-Principe	239	00		7 chiffres	UTC	
Sénégal	221	00		9 chiffres	UTC	
Serbie	381	00	0	4 à 12 chiffres	+1/+2	
Service de coût partagé international (ISCS)	808			8 chiffres		
Service kiosque international (IPRS)	979			9 chiffres		
Service mobile mondial par satellite (GMSS), indicatif partagé	881					
Seychelles	248	00		7 chiffres	+4	
Sierra Leone	232	00	0	8 chiffres	UTC	
Singapour	65	001,008		8 à 12 chiffres	+8	9
Sint Maarten (partie neérlandaise)	1	011	1	(721) + 7 chiffres	-4	
Slovaquie	421	00	0	4 à 9 chiffres	+1/+2	
Slovénie	386	00	0	8 chiffres	+1/+2	
Somalie	252	00		5 à 8 chiffres	+3	
Soudan	249	00	0	9 chiffres	+3	
Soudan du Sud	211	00	0		+3	
Sri Lanka	94	00	0	9 chiffres	+5.30	
Sudafricaine (Rép.)	27	00	0	9 chiffres	+2	2
Suède	46	00	0	7 à 13 chiffres	+1/+2	
Suisse	41	00	0	4 à 12 chiffres	+1/+2	2

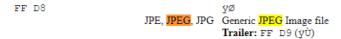
Format File signature:

Hexed.it (http://Hexed.it) (éditeur héxadécimal) :

"JPEG":



GCK'S FILE SIGNATURES TABLE:



Hexed.it (http://Hexed.it) (éditeur héxadécimal) :

La signature est une séquence spécifique d'octets située au début du fichier qui permet d'identifier son format et sa structure.

Une sorte d'étiquette d'identification essentielle qui permet aux logiciels de reconnaître et de traiter les fichiers de manière appropriée.

Source image:

(<u>https://www.itu.int/dms_pub/itu-t/opb/sp/T-SP-E.164C-2011-PDF-F.pdf</u> (https://www.itu.int/dms_pub/itu-t/opb/sp/T-SP-E.164C-2011-

PDF-F.pdf)

(<u>https://www.garykessler.net/library/file_sigs.html</u>

(https://www.garykessler.net/library/file_sigs.html)

Prompt Chat GPT:

"Peux tu me faire un résumé téchnique sur la signature des fichier"

Table ASCII (American Standard Code for Information Interchange)

ASCII TABLE Decimal Hex Decimal Hex Char | Decimal Hex Char | Decimal Hex Char 32 33 34 35 21 22 23 41 42 43 ISTART OF HEADING [START OF HEADING] [START OF TEXT] [END OF TEXT] [END OF TRANSMISSION] c d 68 69 70 71 D 36 24 100 64 [ENOUIRY] 101 38 39 40 41 [ACKNOWLEDGE] 72 73 [HORIZONTAL TAB] 29 49 105 69 [LINE FEED] [VERTICAL TAB] [FORM FEED] [CARRIAGE RETURN] 10 11 12 13 14 15 16 17 18 4D ISHIFT OUT 46 2E 110 78 79 80 81 82 [SHIFT IN] [DATA LINK ESCAPE] [DEVICE CONTROL 1] [DEVICE CONTROL 2] 2F 30 31 111 112 113 47 48 49 50 51 114 [DEVICE CONTROL 2] [DEVICE CONTROL 3] [DEVICE CONTROL 4] [NEGATIVE ACKNOWLEDGE] [SYNCHRONOUS IDLE] 19 20 21 22 23 51 52 53 54 33 83 84 85 86 87 88 89 90 91 92 56 118 IEND OF TRANS BLOCKS 24 25 26 27 56 57 58 59 [CANCEL] [END OF MEDIUM] [SUBSTITUTE] 59 5A 121 122 [ESCAPE] 3B 5B 123 7B 7C [FII F SEPARATOR] [GROUP SEPARATOR] [RECORD SEPARATOR] [UNIT SEPARATOR] 126 127 [DEL]

Définition : La table ASCII est utilisée pour représenter du texte en utilisant des nombres binaires (bits) et est la base de nombreuses autres normes de codage de caractères, notamment UTF-8 et UTF-16, qui prennent en charge un ensemble beaucoup plus large de caractères, y compris les caractères non latins et les emojis.

Source image:

 $(\underline{https://simple.m.wikipedia.org/wiki/File:ASCII-Table-}$

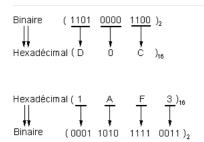
wide.svg (https://simple.m.wikipedia.org/wiki/File:ASCII-Table-wide.svg)

(https://www.garykessler.net/library/file_sigs.html

(https://www.garykessler.net/library/file_sigs.html)

Prompt chat GPT: "C'est quoi une Table ASCII?"

Conversion



Base 10 (Décimal)

• **Définition**: Dans le système décimal, il y a dix chiffres différents disponibles, de 0 à 9. Système de numération le plus couramment utilisé dans le monde.

• Conversions:

- 1. Pour les nombres binaires : Additionnez les puissances de 2 en utilisant les chiffres binaires (0 et 1).
- 2. Pour les nombres hexadécimaux : Utilisez la valeur décimale équivalente de chaque chiffre hexadécimal (0-9, A-F).

Base 2 (Binaire)

• **Définition**: Dans le système binaire, il n'y a que deux chiffres différents disponibles, 0 et 1.

Conversions:

- 1. Pour les nombres décimaux : Divisez successivement par 2, en notant les restes à chaque étape (de droite à gauche) pour obtenir la représentation binaire.
- 2. Pour les nombres hexadécimaux : Convertissez d'abord en base décimale, puis en base binaire.

Le système hexadécimal

 Définition: utilise seize chiffres différents, de 0 à 9 et de A à F (représentant les valeurs de 10 à 15 en décimal)

• Conversions:

- 1. Pour les nombres décimaux : Divisez successivement par 16, en notant les restes à chaque étape, puis lisez les restes de droite à gauche pour obtenir la représentation hexadécimale.
- 2. Pour les nombres binaires : Divisez les chiffres binaires en groupes de 4 (en ajoutant des zéros à gauche si nécessaire) et convertissez chaque groupe en sa valeur hexadécimale équivalente.

il existe plusieurs systèmes de numération différents, mais les systèmes les plus utilisés sont le décimal (base 10), le binaire (base 2) et l'hexadécimal (base 16). Source image: https://www.positron-libre.com/cours/electronique/systeme-numeration/conversion-decimal-binaire-numeration/conversion-decimal-binaire-numeration/conversion-decimal-binaire-numeration/conversion-decimal-binaire-numeration/conversion-decimal-binaire-numeration/conversion-decimal-binaire-numeration/conversion-decimal-binaire-numeration/conversion-decimal-binaire-numeration/conversion-decimal-binaire-numeration/conversion-decimal-binaire-numeration/conversion-decimal-binaire-numeration/conversion-decimal-binaire-numeration/conversion-decimal-binaire-numeration/conversion-decimal-binaire-numeration/conversion-decimal-binaire-numeration/conversion-decimal-binaire-numeration/conversion-decimal-binaire-numeration-decimal-binaire-numera

hexadecimal.php (https://www.positron-libre.com/cours/electronique/systeme-

numeration/conversion-decimal-binaire-hexadecimal.php)

Source text:

https://www.positron-

<u>libre.com/cours/electronique/systeme-</u>

numeration/conversion-decimal-binaire-

hexadecimal.php (https://www.positron-libre.com/cours/electronique/systeme-

numeration/conversion-decimal-binaire-hexadecimal.php)

Prompt Chat GPT : "Que peux-tu me dire à propos de la conversion en base décimale, binaire et hexadécimale? C'est pour un glossaire/aide mémoire technique.""

Programme en Python généré par "chat GPT" pour faire un fichier de dates (iso 8601) to (format csv) :

CSV (Comma-Separated Values) = format de fichier spécifique où les données sont séparées par des virgules

ISO 8601 = un format qui normalise la représentation des dates et heures en utilisant des formats tels que "AAAA-MM-JJ" pour les dates et "hh:mm:ss" pour les heures.

Code phyton générant liste aléatoire de date au format ISO 8601 puis exporté au format CSV :

```
import random
import csv
from datetime import datetime, timedelta

# Function to generate a random ISO 8001 date
lusage
def random_iso8001_date():

start_date = datetime( year 2000, month 1, day 1) # You can adjust the start date as needed
end_date = datetime( year 2023, month 12, day: 31) # You can adjust the end date as needed
random_days = random.randint( a 0, (end_date - start_date).days)
random_date = start_date + timedelta(days=random_days)
return random_date.isoformat()

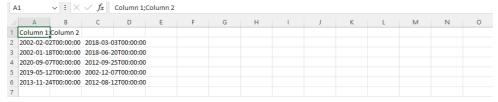
# Generate a list of random ISO 8001 dates
num_dates = 10 # Change this to the number of dates you want to generate
random_dates = [random_iso8001_date() for _ in range(num_dates)]

# Export the data to a CSV file with semicolons as delimiters and a space between the two columns
with open('C:/Users/logod/Desktop/EPSIC/module_102/dates_python/random_dates.csv', 'w', newline='') as csvfile:

csv_writer = csv.writer(csvfile, delimiter=';')
csv_writer.writerow(['Golumn 1:Column 2:']) # Header row
for date1, date2 in zip(random_dates[:num_dates // 2], random_dates[num_dates // 2:]):
csv_writer.writerow([f'{date1} {date2}])

with open('C:/Users/logod/Deskt...
```

Resultat du code effectué dans Excel:



Prompt Chat GPT:

"Yo! Can you help me make a python code for generating a random list of dates displayed in two column with the ISO 8601 format. Generate a different ISO 8601 format for each date.

Then, export all into a CSV file and replace "," by ";" ?"

Changer le titre d'un album, le nom de l'artiste dans un fichier .mp3 à l'aide d'un éditeur hexadécimal. (https://en.wikipedia.org/wiki/ID3

(https://en.wikipedia.org/wiki/ID3)

Définition : Travail en cours **Source :** (lien_de_la_source)

Image 2

Changer le titre d'un album, le nom de l'artiste dans un fichier .mp3 à l'aide d'un éditeur hexadécimal.

EXCEL format des dates

Format d'un mail (RFC 5321)

Format d'une url (RFC 3986)

https://fr.wikipedia.org/wiki/Uniform Resource_Locator)

Locator (https://fr.wikipedia.org/wiki/Uniform_Resource_Locator)

Qu'est ce que la collation du point de vue des caractères en informatique

Définition: Travail en cours

Dans une base de données, la collation est souvent associée à un jeu de caractères (charset) qui définit l'ensemble des caractères pouvant être utilisés dans la base de données

DNS

Qubit vs Bit

bit = 0 ou 1

Quantique

Dans une base de données, la collation est souvent associée à un jeu de caractères (charset) qui définit l'ensemble des caractères pouvant être utilisés dans la base de données

4

. Par exemple, le jeu de caractères UTF-8 permet de représenter tous les caractères Unicode, tandis que le jeu de caractères ASCII ne permet de représenter que les caractères de base de l'alphabet latin 3

le débit ≠ vitesse

Entité

N'importe quoi qui peut être représenté dans un ordinateur