

TD 1a — Fichiers, formats et extensions

Guide détaillé, étape par étape

Le TD manipule les fichiers d'un même texte, le poème *The Raven* d'Edgar Allan Poe (1845), sous plusieurs formes : un document LibreOffice, un fichier texte, deux pages web. On les exporte, on les copie, on change leur extension et on les ouvre avec différents logiciels, pour comparer ce que décide l'extension et ce que contient le fichier. Il dure une vingtaine de minutes.

Tout se fait avec l'explorateur de fichiers, LibreOffice, un navigateur et deux éditeurs de texte, le Bloc-notes et Notepad++. Aucune commande n'est nécessaire.

Le dossier `depart\` contient aussi les mêmes fichiers pour un second texte, *Auld Lang Syne* de Robert Burns (1788). Qui a fini en avance peut refaire les étapes avec lui.

Étape	Objectif	Ce qu'on fait
1	travailler sur des copies, à côté des fichiers livrés	préparer le dossier du TD
2	produire un même texte dans plusieurs formats	exporter un même document en trois formats
3	tester ce que l'extension change à l'ouverture	changer l'extension de copies du document
4	lire l'adresse d'une page ouverte depuis le disque	ouvrir une page web depuis son disque
5	comparer fichiers texte et fichiers binaires dans un éditeur	lire les fichiers avec deux éditeurs de texte

Chaque étape commence par un encadré qui la résume. Ce que chaque étape fait constater est expliqué à la fin du guide, dans « Ce que le TD fait constater » : faire l'étape d'abord, et noter ce qu'on observe, avant de lire l'explication.

1 · Préparer le dossier du TD

À faire : récupérer l'archive de la séance ; afficher les extensions dans l'explorateur ; ouvrir le dossier `cours1\1a_formats\`.

À obtenir : l'explorateur montre `depart\` et `travail\`, et les noms de fichiers se terminent par leur extension (`raven.odt`, pas `raven`).

Récupérer les fichiers

Copier l'archive `info01-cours1.zip` depuis le dossier partagé `formationTemp` dans le dossier `info01` du Bureau, puis l'extraire dans ce même dossier (clic droit, « Extraire tout... », en effaçant la fin du dossier proposé). La page « Récupérer les fichiers d'une séance » du site du cours détaille ces deux opérations. Ne pas travailler dans le dossier partagé.

Afficher les extensions

Windows masque par défaut les extensions des types de fichiers qu'il connaît. Dans l'explorateur : menu Affichage, Afficher, cocher « Extensions de noms de fichiers ». Sous macOS : Finder, Réglages, Avancé, « Afficher tous les suffixes de fichiers ». Le réglage se fait une fois, et sert tout le semestre.

Les deux dossiers du TD

Ouvrir `info01\cours1\1a_formats\`. Le dossier contient :

```
1a_formats\  
├─ depart\  
│   ├─ raven.odt           le poème, document LibreOffice  
│   ├─ raven_une_ligne.txt le poème, en texte sur une seule ligne  
│   ├─ raven_une_ligne.donnees le même fichier, avec une autre extension  
│   ├─ raven_brut.html     le poème, en page web sans mise en forme  
│   ├─ raven_style.html   la même page, avec une feuille de style  
│   ├─ style.css           la feuille de style  
│   └─ auld_lang_syne...  les mêmes fichiers pour le second texte  
├─ travail\               vide  
├─ td_1a_formats.pdf      la feuille du TD  
└─ README.md
```

`depart\` contient les fichiers fournis, et ne se modifie pas. `travail\` reçoit les copies et ce que le TD fabrique. Si une copie est abîmée, on en refait une à partir de `depart\`.

Le chemin complet d'un fichier de départ, et celui d'une copie, sont :

```
C:\Users\eleve\Desktop\info01\cours1\1a_formats\depart\raven.odt  
C:\Users\eleve\Desktop\info01\cours1\1a_formats\travail\raven.odt.pdf
```

La suite du guide écrit les chemins à partir de `1a_formats\`.

Vérification : la barre d'adresse de l'explorateur se termine par `cours1\1a_formats` (ou « Bureau > info01 > cours1 > 1a_formats »), et `depart\` montre `raven.odt` avec son extension.

2 · Exporter un même document en trois formats

À faire : ouvrir `depart\raven.odt` dans LibreOffice Writer ; l'exporter en PDF puis en PNG dans `travail\` ; rouvrir les trois fichiers.

À obtenir : `travail\` contient `raven.pdf` et `raven.png`.

Double-cliquer sur `depart\raven.odt` : LibreOffice Writer l'ouvre.

1. Menu Fichier, Exporter au format PDF, puis Exporter. Enregistrer dans `travail\` sous le nom `raven.pdf`.
2. Menu Fichier, Exporter..., choisir le type PNG. Enregistrer dans `travail\` sous le nom `raven.png`. L'export en image est dans « Exporter », pas dans « Enregistrer sous ».
3. Fermer LibreOffice sans enregistrer le `.odt`.

Rouvrir ensuite les trois fichiers, et essayer dans chacun de sélectionner une ligne du poème, puis de chercher un mot avec `Ctrl + F`. Le `.png` s'ouvre dans une visionneuse d'images ; ouvert depuis LibreOffice, il s'affiche dans Draw, comme une image posée sur une page.

À noter : pour chacun des trois fichiers, si le texte est encore du texte (il se sélectionne, se cherche, se modifie).

3 · Changer l'extension de copies du document

À faire : faire quatre copies de `depart\raven.odt` dans `travail\`, sous quatre noms différents ; les ouvrir par double-clic.

À obtenir : `travail\` contient `raven_odt.pdf`, `raven_odt.jpg`, `riri.fifi.loulou.odt` et `raven.loulou`.

Copier et renommer

Dans l'explorateur :

1. sélectionner `depart\raven.odt`, puis `Ctrl + C` ;
2. ouvrir `travail\`, puis `Ctrl + V` ;
3. sélectionner la copie, `F2`, taper le nouveau nom, puis `Entrée` ;
4. Windows demande de confirmer le changement d'extension : répondre Oui.

Qui connaît déjà le terminal peut faire la même chose au clavier. Dans l'explorateur, clic droit sur le dossier `la_formats`, « Ouvrir dans le terminal », puis, dans un terminal `cmd` :

```
copy depart\raven.odt travail\raven_odt.pdf
copy depart\raven.odt travail\raven.loulou
```

Sous macOS et Linux, la commande est `cp depart/raven.odt travail/raven_odt.pdf`. Le terminal n'est pas nécessaire au TD ; le cours 2 lui consacre une partie.

Les quatre copies

Faire les quatre copies, puis double-cliquer sur chacune :

Nom de la copie	Ce que le système lance
<code>raven_odt.pdf</code>	un lecteur PDF
<code>raven_odt.jpg</code>	une visionneuse d'images
<code>riri.fifi.loulou.odt</code>	LibreOffice Writer
<code>raven.loulou</code>	rien : Windows demande quel logiciel choisir ; choisir LibreOffice Writer

À noter : pour chaque copie, si elle s'ouvre, et le message affiché quand elle ne s'ouvre pas. La visionneuse d'images en donne un précis.

4 · Ouvrir une page web depuis son disque

À faire : ouvrir `depart\raven_brut.html` puis `depart\raven_style.html` dans le navigateur ; modifier une couleur dans `style.css` ; lire l'adresse de la page ; ouvrir une copie dont le nom contient un espace.

À obtenir : la page mise en forme change de couleur après F5 ; `travail\` contient `raven style.html`.

Deux pages, une feuille de style

1. Double-cliquer sur `depart\raven_brut.html`. Le navigateur l'ouvre, sans réseau.
2. Double-cliquer sur `depart\raven_style.html` : le même texte, mis en forme.
3. Ouvrir les deux fichiers dans le Bloc-notes (clic droit, Ouvrir avec, Bloc-notes) et comparer leur en-tête, entre `<head>` et `</head>`, puis la façon dont le poème est écrit.
4. Ouvrir `depart\style.css` dans le Bloc-notes. Remplacer la valeur d'une ligne `color:` par `crimson`, ou par un code comme `#c0392b`, et enregistrer.
5. Revenir au navigateur sur `raven_style.html`, et recharger la page avec F5.

`style.css` est le seul fichier de `depart\` que le TD modifie : remettre sa couleur d'origine à la fin de l'étape.

L'adresse de la page

Lire l'adresse que le navigateur affiche pour `raven_style.html`. Elle a la forme :

```
file:///C:/Users/eleve/Desktop/info01/cours1/1a_formats/depart/raven_style.html
```

À noter : ce qui ressemble à un chemin de fichier dans cette adresse, et ce qui en diffère. Noter aussi ce qui distingue les deux fichiers HTML dans le Bloc-notes.

Un espace dans le nom d'un fichier

1. Copier `depart\raven_style.html` dans `travail\`, et renommer la copie `raven style.html`, avec un espace.
2. L'ouvrir par double-clic, et lire l'adresse affichée.
3. Copier l'adresse (clic dans la barre, Ctrl + A, Ctrl + C), et la coller dans le Bloc-notes.

À noter : l'apparence de la page, comparée à `depart\raven_style.html` ; comment l'espace apparaît dans la barre d'adresse, puis dans l'adresse collée. Le résultat dépend du navigateur.

5 · Lire les fichiers avec deux éditeurs de texte

À faire : ouvrir quatre fichiers de depart\ dans le Bloc-notes, puis dans Notepad++ ; retrouver des caractères dans la table ASCII.

À obtenir : les quatre fichiers ouverts dans chaque éditeur, puis fermés sans enregistrer.

La table ASCII

Un fichier texte range chaque caractère dans un octet, selon une table. La table ASCII en donne 95, avec leur valeur en hexadécimal : la ligne donne le premier chiffre, la colonne le second. R est à la ligne 5_ et dans la colonne 2 : il vaut 52.

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
2_	␣	!	"	#	\$	%	&	'	(	)	*	+	,	-	.	/
3_	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4_	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5_	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
6_	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7_	p	q	r	s	t	u	v	w	x	y	z	{		}	~	

␣ représente l'espace. Les valeurs de 00 à 1F ne sont pas des caractères affichables mais des commandes, dont le saut de ligne, 0A. 7F non plus.

Retrouver dans la table la valeur de l'espace, puis celle de A et celle de a.

À noter : l'écart entre la valeur d'une majuscule et celle de la minuscule correspondante.

Deux éditeurs, les mêmes fichiers

Ouvrir chacun des quatre fichiers suivants dans le Bloc-notes (clic droit, Ouvrir avec, Bloc-notes), puis dans Notepad++ (clic droit, Ouvrir avec, Notepad++, ou par le menu Fichier de Notepad++) :

Fichier, dans depart\	Dans le Bloc-notes	Dans Notepad++
raven_une_ligne.txt		
style.css		
raven_brut.html		
raven.odt		

Attention : ne rien enregistrer, et fermer sans sauver. Un .odt réenregistré par un éditeur de texte est détruit.

À noter : pour chaque fichier et chaque éditeur, si le contenu est lisible, et ce que Notepad++ ajoute à l'affichage.

Ce que le TD fait constater

Cette section se lit après avoir fait les étapes.

Étape 2 : trois formats d'un même document

Fichier	Le texte est-il encore du texte ?
raven.odt	oui, et il reste modifiable dans LibreOffice
raven.pdf	oui : il se sélectionne et se cherche, mais la mise en page est figée
raven.png	non : ce sont des pixels, et seule la première page est exportée

Un PDF décrit la page : les caractères y sont rangés avec leur position. Une image PNG ou JPEG photographie la page : elle ne garde que l'apparence, et le texte n'y est plus que des points de couleur.

Étape 3 : l'extension décide du logiciel, pas du contenu

Nom de la copie	Ce qui se passe
raven_odt.pdf	refus : le lecteur PDF ne reconnaît pas le fichier
raven_odt.jpg	refus : <i>Not a JPEG file: starts with 0x50 0x4b</i>
riri.fifi.loulou.odt	s'ouvre : seule la fin du nom, après le dernier point, compte
raven.loulou	s'ouvre une fois LibreOffice choisi : le contenu n'a pas changé

Le système ne regarde que le nom. Avec une extension qu'il connaît, il lance le logiciel associé, qui refuse le fichier s'il n'est pas du format attendu. Avec une extension inconnue, Windows et macOS demandent quel logiciel employer, et retiennent ce choix si on coche « Toujours ». Linux, lui, regarde les premiers octets du fichier et propose LibreOffice de lui-même.

0x50 0x4b sont les deux premiers octets du fichier, écrits en hexadécimal. Dans la table ASCII, ce sont les caractères P et K : la signature d'une archive ZIP, dont un .odt est une forme (TD 1b).

Étape 4 : une page web, son style, son adresse

raven_brut.html et raven_style.html contiennent le même poème. Dans raven_brut.html, il tient dans un seul paragraphe, <p>, et le navigateur l'affiche en un bloc : les sauts de ligne du fichier ne sont pas rendus, la structure d'une page HTML se déclare par des balises. raven_style.html écrit un paragraphe par vers, et son en-tête contient une ligne de plus, <link rel="stylesheet" href="style.css">, qui appelle la feuille de style. style.css décrit la présentation : couleurs, polices, largeur. Modifier style.css change l'apparence de la page sans toucher au texte. Le contenu est dans un fichier, la présentation dans un autre, et l'un change sans l'autre. CSS accepte les couleurs par leur nom (crimson) comme par leur code hexadécimal (#c0392b), deux chiffres par composante, rouge, vert, bleu.

Une adresse web, ou **URL**, est un chemin de fichier précédé de la machine où aller le chercher :

Partie	Dans https://www.ensg.eu/cours/info01/raven.html
le protocole, la façon convenue de demander le fichier	https://
la machine	www.ensg.eu
le chemin sur cette machine	/cours/info01/
le fichier	raven.html

L'adresse d'une page ouverte depuis le disque a la même forme, avec le protocole `file`. Entre `file://` et le chemin, la place de la machine est vide, puisque c'est la vôtre ; la troisième barre est le début du chemin, `/C:/`. Le navigateur écrit tous les chemins avec des `/`, même sous Windows.

Une adresse ne peut pas contenir d'espace : il y est écrit `%20`, `20` étant la valeur hexadécimale de l'espace dans la table ASCII. Chrome et Edge affichent `raven%20style.html` dans la barre ; Firefox affiche l'espace, mais l'adresse copiée contient `%20` dans tous les cas. Ce que la barre affiche est une présentation, ce qui se copie est l'adresse réelle.

La copie `raven style.html` s'affiche sans mise en forme, alors que l'original de `depart\` en a une. `href="style.css"` est un chemin relatif : il désigne `style.css` dans le dossier de la page. La copie est dans `travail\`, où il n'y a pas de `style.css`, et le navigateur ne trouve pas la feuille de style.

Un espace dans un nom de fichier passe dans l'explorateur et dans le navigateur, mais il demande des guillemets en ligne de commande, au cours 3.

Étape 5 : ce qu'un éditeur de texte affiche

La table ASCII donne `20` pour l'espace, `41` pour `A` et `61` pour `a` : une minuscule vaut sa majuscule plus `20`.

Fichier	Dans le Bloc-notes	Dans Notepad++
<code>raven_une_ligne.txt</code>	le poème, lisible en entier	le même, sans couleur
<code>style.css</code>	des règles, lisibles	sélecteurs et propriétés en couleur
<code>raven_brut.html</code>	le texte et ses balises	les balises en couleur, repliables
<code>raven.odt</code>	PK, puis des caractères sans suite : du binaire	les mêmes, et des NUL en surbrillance

Un éditeur de texte affiche un caractère par octet, selon la table d'encodage, et n'interprète rien d'autre : ni image, ni mise en forme. Ce qui n'a pas de caractère correspondant apparaît en carré ou en signe étrange ; Notepad++ marque NUL les octets de valeur `00`.

Les couleurs de Notepad++ ne viennent pas du fichier : l'éditeur les ajoute d'après l'extension, qu'il associe à un langage. Renommer `style.css` en `style.txt` les fait disparaître. L'éditeur de code de la partie 2 fait la même chose pour Python.

`raven_brut.html` est du texte, `raven.odt` n'en est pas, et leur nom ne le disait pas : seule l'ouverture dans un éditeur de texte le montre. Les PK du début sont la signature ZIP rencontrée à l'étape 3.