



виявлення категорій даних у документах

Редакція цього файлу:

- містить демонстрацію можливого способу вирішення поставленої задачі
- оновлена 04 листопада 2019 (версія 1.0.1)
- розміщена на <https://legaltech.org.ua/cases>

Ось 3 демонстраційні документи:

документ 1

Я, Заєць Павло Григорович, даю Сокирі Петру Сергійовичу 2000 гривень у борг. Дата передачі грошей – 02 серпня 2019 року. Строк для повернення – 10 днів. Відсотків нема.

документ 2

Я, Їжак Ганна Василівна, 04.08.2019 даю Квітці Ірині Степанівні 3000 грн. у борг строком на 12 днів, без відсотків.

документ 3

15 квітня 2019 я, Коваль А. П., даю Романенко Катерині Павлівні 100 доларів США на 18 днів, повернути треба на 2 відсотки більше.

Починаємо виявляти категорії даних у конкретних їх фрагментах:

документ 1

Я, Заєць Павло Григорович, даю Сокирі Петру Сергійовичу 2000 гривень у борг. Дата передачі грошей – 02 серпня 2019 року. Строк для повернення – 10 днів. Відсотків нема.

документ 2

Я, Їжак Ганна Василівна, 04.08.2019 даю Квітці Ірині Степанівні 3000 грн. у борг строком на 12 днів, без відсотків.

документ 3

15 квітня 2019 я, Коваль А. П., даю Романенко Катерині Павлівні 100 доларів США на 18 днів, повернути треба на 2 відсотки більше.

фрагмент/категорія даних – "кредитор":

документ 1

Я, **Заєць Павло Григорович**, даю Сокирі Петру Сергійовичу 2000 гривень у борг. Дата передачі грошей – 02 серпня 2019 року. Строк для повернення – 10 днів. Відсотків нема.

документ 2

Я, **Їжак Ганна Василівна**, 04.08.2019 даю Квітці Ірині Степанівні 3000 грн. у борг строком на 12 днів, без відсотків.

документ 3

15 квітня 2019 я, **Коваль А. П.**, даю Романенко Катерині Павлівні 100 доларів США на 18 днів, повернути треба на 2 відсотки більше.

фрагмент/категорія даних – "боржник":

документ 1

Я, **Заєць Павло Григорович**, даю **Сокирі Петру Сергійовичу** 2000 гривень у борг. Дата передачі грошей – 02 серпня 2019 року. Строк для повернення – 10 днів. Відсотків нема.

документ 2

Я, **Їжак Ганна Василівна**, 04.08.2019 даю **Квітці Ірині Степанівні** 3000 грн. у борг строком на 12 днів, без відсотків.

документ 3

15 квітня 2019 я, **Коваль А. П.**, даю **Романенко Катерині Павлівні** 100 доларів США на 18 днів, повернути треба на 2 відсотки більше.

фрагмент/категорія даних – "сума позики":

документ 1

Я, **Заєць Павло Григорович**, даю **Сокирі Петру Сергійовичу** **2000 гривень** у борг. Дата передачі грошей – 02 серпня 2019 року. Строк для повернення – 10 днів. Відсотків нема.

документ 2

Я, **Їжак Ганна Василівна**, 04.08.2019 даю **Квітці Ірині Степанівні** **3000 грн.** у борг строком на 12 днів, без відсотків.

документ 3

15 квітня 2019 я, **Коваль А. П.**, даю **Романенко Катерині Павлівні** **100 доларів США** на 18 днів, повернути треба на 2 відсотки більше.

фрагмент/категорія даних – "дата початку зобов'язань":

документ 1

Я, Заєць Павло Григорович, даю Сокирі Петру Сергійовичу 2000 гривень у борг. Дата передачі грошей – 02 серпня 2019 року. Строк для повернення – 10 днів. Відсотків нема.

документ 2

Я, Іжак Ганна Василівна, 04.08.2019 даю Квітці Ірині Степанівні 3000 грн. у борг строком на 12 днів, без відсотків.

документ 3

15 квітня 2019 я, Коваль А. П., даю Романенко Катерині Павлівні 100 доларів США на 18 днів, повернути треба на 2 відсотки більше.

фрагмент/категорія даних – "строк зобов'язань":

документ 1

Я, Заєць Павло Григорович, даю Сокирі Петру Сергійовичу 2000 гривень у борг. Дата передачі грошей – 02 серпня 2019 року. Строк для повернення – 10 днів. Відсотків нема.

документ 2

Я, Іжак Ганна Василівна, 04.08.2019 даю Квітці Ірині Степанівні 3000 грн. у борг строком на 12 днів, без відсотків.

документ 3

15 квітня 2019 я, Коваль А. П., даю Романенко Катерині Павлівні 100 доларів США на 18 днів, повернути треба на 2 відсотки більше.

фрагмент/категорія даних – "чи є відсотки, і якщо є, то скільки":

документ 1

Я, Заєць Павло Григорович, даю Сокирі Петру Сергійовичу 2000 гривень у борг. Дата передачі грошей – 02 серпня 2019 року. Строк для повернення – 10 днів. Відсотків нема.

документ 2

Я, Іжак Ганна Василівна, 04.08.2019 даю Квітці Ірині Степанівні 3000 грн. у борг строком на 12 днів, без відсотків.

документ 3

15 квітня 2019 я, Коваль А. П., даю Романенко Катерині Павлівні 100 доларів США на 18 днів, повернути треба на 2 відсотки більше.

Отже, в нас є:

- документи (кейси, ситуації)
- категорії даних по кожному документу (кейсу, ситуації)

Отже, в нас є:

- документи (кейси, ситуації)
- категорії даних по кожному документу (кейсу, ситуації)

категорії даних

Документи

Це утворює таблицю, що є способом організації реляційних баз даних

id	кредитор	боржник	сума	дата, від	днів	відсотки
1	Заєць Павло Григорович	Сокирі Петру Сергійовичу	2000 гривень	02 серпня 2019 року	10	0
2	Іжак Ганна Василівна	Квітці Ірині Степанівні	3000 гривень	04.08.2019	12	0
3	Коваль А. П.	Романенко Катерині Павлівні	100 доларів США	15 квітня 2019	18	2

Тепер необхідно оцінити повноту представлення даних

id	кредитор	боржник	сума	дата, від	днів	відсотки
1	Заєць Павло Григорович	Сокирі Петру Сергійовичу	2000 гривень	02 серпня 2019 року	10	0
2	Іжак Ганна Василівна	Квітці Ірині Степанівні	3000 гривень	04.08.2019	12	0
3	Коваль А. П.	Романенко Катерині Павлівні	100 доларів США	15 квітня 2019	18	2

Тепер необхідно оцінити повноту представлення даних

id	кредитор	боржник	сума	дата, від	днів	відсотки
1	Заєць Павло Григорович	Сокирі Петру Сергійовичу	2000 гривень	02 серпня 2019 року	10	0
2	Іжак Ганна Василівна	Квітці Ірині Степанівні	3000 гривень	04.08.2019	12	0
3	Коваль А. П.	Романенко Катерині Павлівні	100 доларів США	15 квітня 2019	18	2

Проблема 1 – неповнота даних.

(мається на увазі неповний запис імені та по-батькові)

Тепер необхідно оцінити повноту представлення даних

id	кредитор	боржник	сума	дата, від	днів	відсотки
1	Заєць Павло Григорович	Сокирі Петру Сергійовичу	2000 гривень	02 серпня 2019 року	10	0
2	Їжак Ганна Василівна	Квітці Ірині Степанівні	3000 гривень	04.08.2019	12	0
3	Коваль А. П.	Романенко Катерині Павлівні	100 доларів США	15 квітня 2019	18	2

Проблема 2 – незручна форма представлення даних.

(мається на увазі давальний відмінок)

Тепер необхідно оцінити однорідність представлення даних

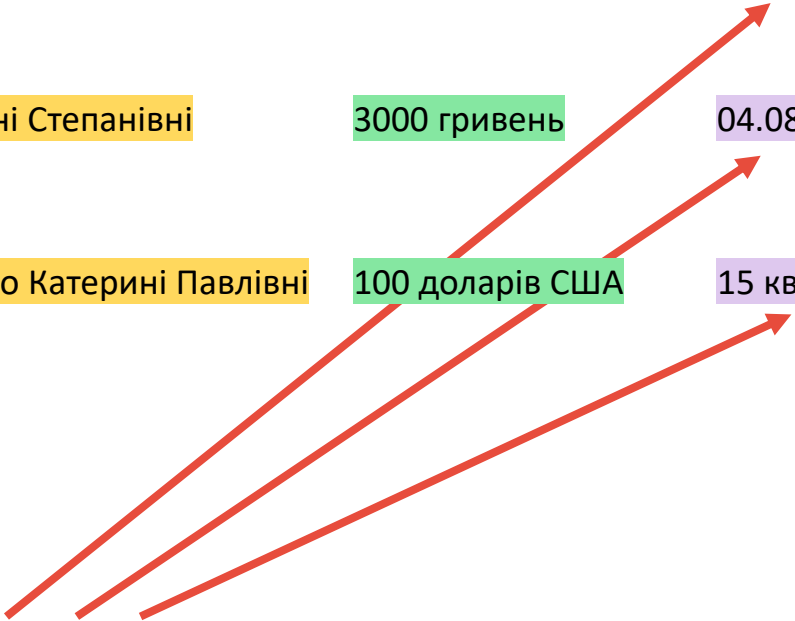
id	кредитор	боржник	сума	дата, від	днів	відсотки
1	Заєць Павло Григорович	Сокирі Петру Сергійовичу	2000 гривень	02 серпня 2019 року	10	0
2	Іжак Ганна Василівна	Квітці Ірині Степанівні	3000 гривень	04.08.2019	12	0
3	Коваль А. П.	Романенко Катерині Павлівні	100 доларів США	15 квітня 2019	18	2

Проблема 3 – неоднорідність систем відліку даних.

(мається на увазі змішування різних валют)

Тепер необхідно оцінити однорідність представлення даних

id	кредитор	боржник	сума	дата, від	днів	відсотки
1	Заєць Павло Григорович	Сокирі Петру Сергійовичу	2000 гривень	02 серпня 2019 року	10	0
2	Іжак Ганна Василівна	Квітці Ірині Степанівні	3000 гривень	04.08.2019	12	0
3	Коваль А. П.	Романенко Катерині Павлівні	100 доларів США	15 квітня 2019	18	2



Проблема 4 – різність форми запису даних.

(мається на увазі змішування різних валют)

Тепер пропонуємо вирішення проблем

id	кредитор	боржник	сума	дата, від	днів	відсотки
1	Заєць Павло Григорович	Сокирі Петру Сергійовичу	2000 гривень	02 серпня 2019 року	10	0
2	Іжак Ганна Василівна	Квітці Ірині Степанівні	3000 гривень	04.08.2019	12	0
3	Коваль А. П.	Романенко Катерині Павлівні	100 доларів США	15 квітня 2019	18	2

Проблема 1 – неповнота даних.

Рішення: знайти дані. Якщо не можемо, то лишаємо, як є.

Тепер пропонуємо вирішення проблем

id	кредитор	боржник	сума	дата, від	днів	відсотки
1	Заєць Павло Григорович	Сокира Петро Сергійович	2000 гривень	02 серпня 2019 року	10	0
2	Їжак Ганна Василівна	Квітка Ірина Степанівна	3000 гривень	04.08.2019	12	0
3	Коваль А. П.	Романенко Катерина Павлівна	100 доларів США	15 квітня 2019	18	2



Проблема 2 – незручна форма представлення даних.

Рішення: можемо за словником змінити відмінок.

Тепер пропонуємо вирішення проблем

id	кредитор	боржник	сума	дата, від	днів	відсотки
1	Заєць Павло Григорович	Сокира Петро Сергійович	2000 гривень	02 серпня 2019 року	10	0
2	Іжак Ганна Василівна	Квітка Ірина Степанівна	3000 гривень	04.08.2019	12	0
3	Коваль А. П.	Романенко Катерина Павлівна	2655 гривень	15 квітня 2019	18	2


дата X

Проблема 3 – неоднорідність систем відліку даних.

Рішення: зможемо обчислити борг у гривні станом на дату X.

Тепер пропонуємо вирішення проблем

id	кредитор	боржник	сума	дата, від	днів	відсотки
1	Заєць Павло Григорович	Сокира Петро Сергійович	2000 гривень	02.08.2019	10	0
2	Іжак Ганна Василівна	Квітка Ірина Степанівна	3000 гривень	04.08.2019	12	0
3	Коваль А. П.	Романенко Катерина Павлівна	2655 гривень	15.04.2019	18	2



Проблема 4 – різність форми запису даних.

Рішення: зводимо до формату "dd.mm.YYYY".

Трансформація в базу даних змушує типізувати дані

id	кредитор	боржник	сума, грн	дата, від	днів	відсотки
1	Заєць Павло Григорович	Сокира Петро Сергійович	2000	02.08.2019	10	0
2	Їжак Ганна Василівна	Квітка Ірина Степанівна	3000	04.08.2019	12	0
3	Коваль А. П.	Романенко Катерина Павлівна	2655	15.04.2019	18	2
integer	string	string	integer	string or date	integer	integer

Трансформація в базу даних змушує типізувати дані

id	кредитор	боржник	сума, грн	дата, від	днів	відсотки
1	Заєць Павло Григорович	Сокира Петро Сергійович	2000	02.08.2019	10	0
2	Їжак Ганна Василівна	Квітка Ірина Степанівна	3000	04.08.2019	12	0
3	Коваль А. П.	Романенко Катерина Павлівна	2655	15.04.2019	18	2
integer	string	string	integer	string or date	integer	integer
число	рядок	рядок	число	рядок чи дата	число	число



Андрій Костенко

Legal Engineer в Axon Partners



facebook.com/andrkostenko



t.me/andrkosten

axon.partners



t.me/legalcode



<https://legaltech.org.ua/cases>

IP attribution slide

different pictures and icons are made:

- by [Freepik](#) from www.flaticon.com

different pictures and icons are taken from:

- <https://emojipedia.org>