Цель и задачи

Повысить точность сравнения товаров в прайс-листах, оформленных в разных стандартах описаний

Исследовать существующие способы сопоставления товаров

Определение методики сопоставления текстовых данных

Разработка системы сравнения описаний товаров

Введение

Структурированные данные		Неструктурированные данные
Категория	Ноутбук	Ноутбук HP 15-db0080ur <4JW37EA> AMD A9-9425 (3.0)/8Gb/1Tb/15.6""HD AG/AMD 520 2GB/No ODD/Cam HD/Win10 (Scarlet Red)
HP	HP	
ОЗУ	8Gb	
HDD	1Tb	
Операционная система	Win10	
Веб-камера	Cam HD	
Процессор	i7-8750H(2.2)	

Введение

[Hoyтбук] HP Envy x360 15-bp104ur [2pq27ea] silver 15,6" i5-8250U/8Gb/1Tb/128Gb SSD/GF Mx150 4Gb/W10

"15.6"" Ноутбук HP Envy x360 15-bp104ur (FHD/IPS/Touch) i5 8250U(1.6)/8192/1Tb+SSD 128/NV MX150 4Gb/Win10/Silver"

Одинаковые свойства		Одинаковые свойства, но отличающиеся написанием		Дополнительные свойства
Envy x360	Envy x360	[Ноутбук]	Ноутбук	[2pq27ea]
15-bp104ur	15-bp104ur	15,6"	"15.6""	(FHD/IPS/Touch)
		i5-8250U	i5 8250U(1.6)	
		8Gb	8192	
		1Tb/128Gb SSD	1Tb+SSD 128	
		GF Mx150 4Gb	NV MX150 4Gb	
		W10	Win10	

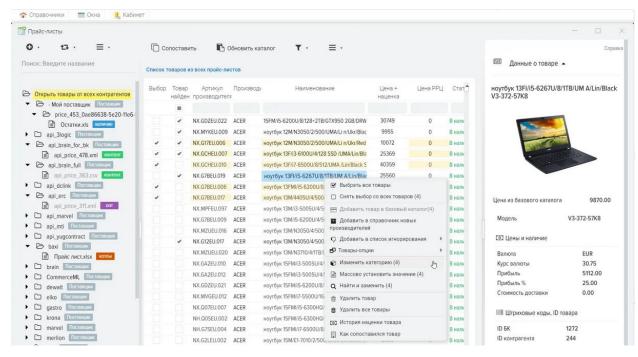
Введение

Hoyтбук HP Pavilion Gaming 17-ab408ur <4GX31EA> i7-8750H (2.2)/8Gb/1TB+128Gb SSD/17.3"" IPS FHD AG/NV GTX 1050Ti 4GB/DVD-RW/Cam HD/Win10 (Shadow black)

Ноутбук HP 15-db0080ur <4JW37EA> AMD A9-9425 (3.0)/8Gb/1Tb/15.6""HD AG/AMD 520 2GB/No ODD/Cam HD/Win10 (Scarlet Red)

Одинаковые свойства		Разные свойства		Дополнительные свойства
Ноутбук	Ноутбук	17-ab408ur	15-db0080ur	Pavilion Gaming
HP	HP	<4GX31EA>	<4JW37EA>	+128Gb SSD
8Gb	8Gb	i7-8750H(2.2)	AMD A9-9425	
1TB	1Tb	17.3""	15.6""	
Win10	Win10	IPS FHD AG	HD AG	
Cam HD	Cam HD	NV GTX 1050Ti 4 GB	AMD 520 2GB	
		(Shadow black)	(Scarlet Red)	

Анализ существующего программного обеспечения



Существующие решения ориентированы контроль товаров в компании и на поиск вхождений, а также на ручную регулировку каждого из рассматриваемых прайслистов.

Таким образом, готовые решения подходят в основном для малого бизнеса, в которых не большая численность товаров.

Крупные компании скорее ориентируются на собственные стандарты описаний, которым должны следовать поставщики, что затрудняет их взаимодействие.

Требования к разрабатываемой системе

- Принимать два объекта строковых данных, в которых содержатся свойства товаров
- Структурировать и классифицировать строковые данные в отдельные свойства, согласно их категории
- Сравнить эти свойства между товарами, к которыми они принадлежат, согласно наименованию свойств
- Определить однозначный результат сравнения двух объектов

Архитектура системы

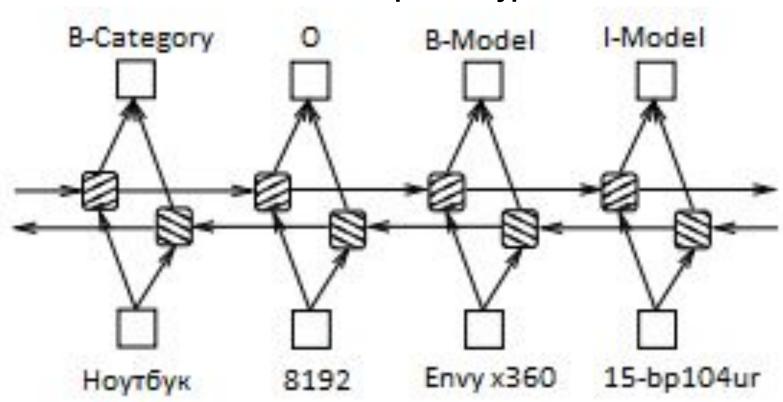
Система будет состоять из следующих модулей:

- UI модуль
- Модуль структуризации данных
- Модуль сопоставления данных



Модуль структуризации

BI-LSTM Архитектура



Именованные сущности

"15.6"" Ноутбук HP Envy x360 15-bp104ur (FHD/IPS/Touch) i5 8250U(1.6)/8192/1Tb+SSD 128/NV MX150 4Gb/Win10/Silver"

Наименование свойства	Значение свойства	
Категория	Ноутбук	
Диагональ экрана	"15.6""	
Модель	Envy x360 15-bp104ur	
Оперативная память (ОЗУ)	8192	
Жесткие диски (ПЗУ)	1Tb+SSD 128	
Видеокарта	NV MX150 4Gb	
Матрица экрана	IPS	
Модель процессора	i5 8250U(1.6)	