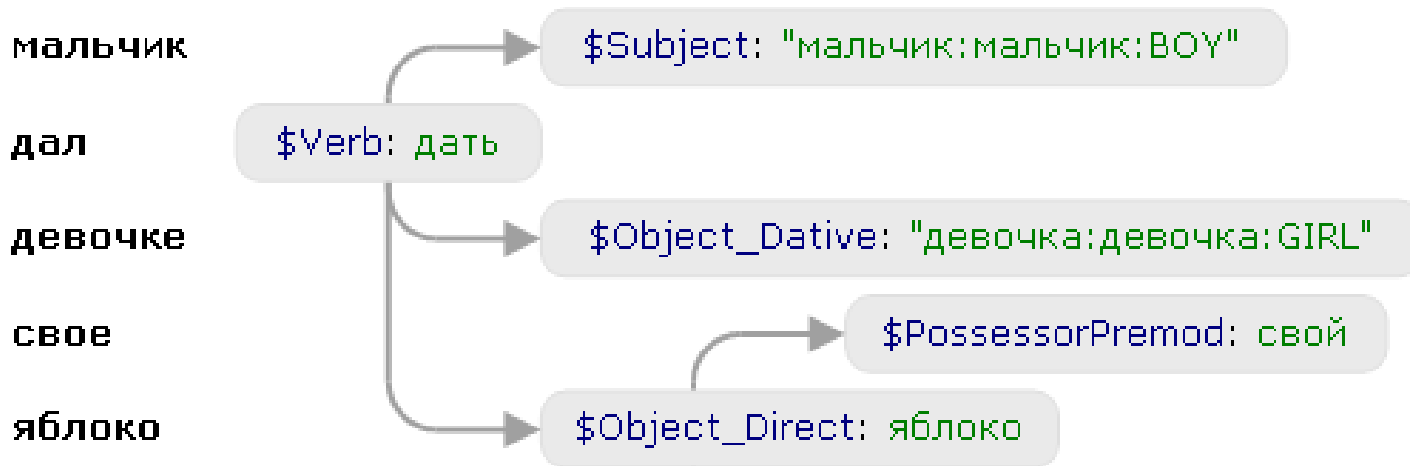


ANAPHORA ANALYSIS BASED ON ABBYY COMPRENO LINGUISTIC TECHNOLOGIES

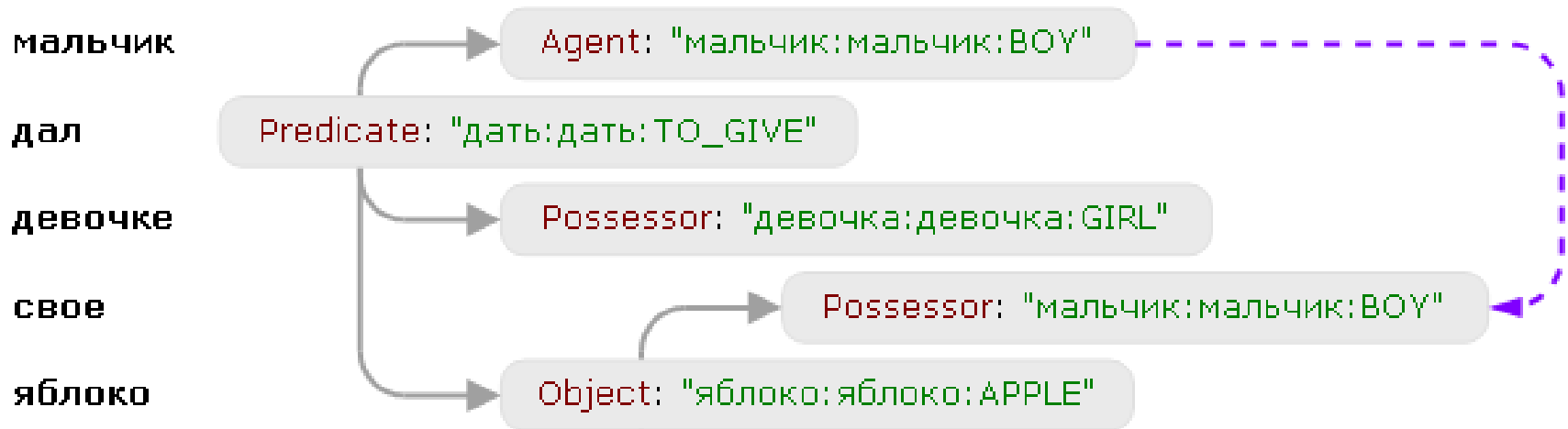
Bogdanov A. V. (abogdanov@abbyy.com)
Dzhumaev S. S. (sdzhumaev@abbyy.com)
Skorinkin D. A. (dskorinkin@abbyy.com)
Starostin A. S. (astarostin@abbyy.com)

Мальчик дал девочке свое яблоко.
 Syntactic tree without non-tree links



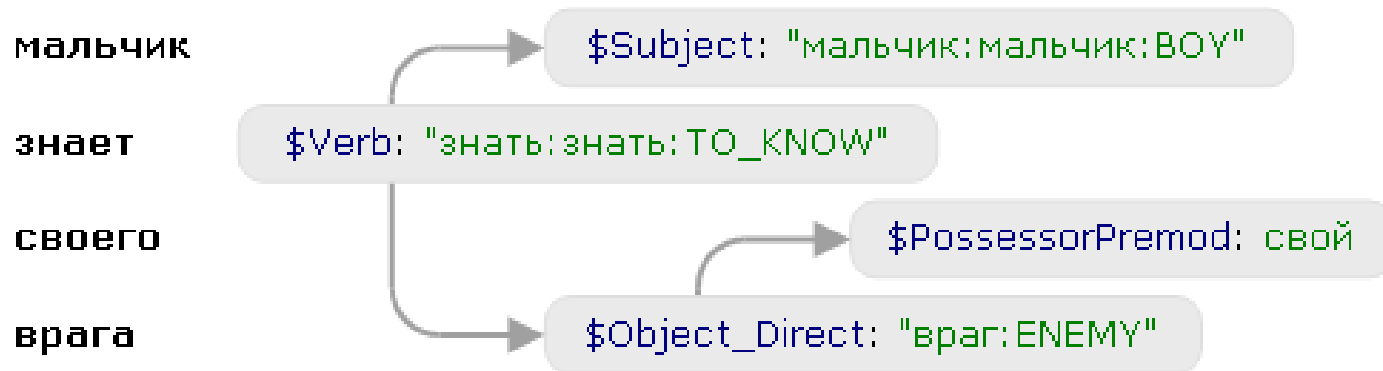
Мальчик дал девочке свое яблоко.

Semantic tree with non-tree links



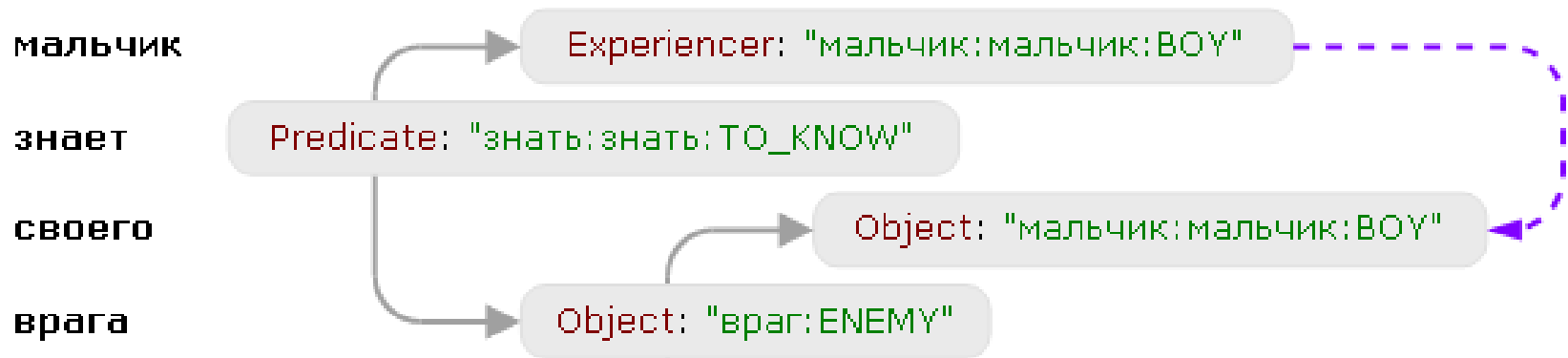
Мальчик знает своего врага.

Syntactic tree without non-tree links



Мальчик знает своего врага.

Semantic tree with non-tree links



Pronoun anaphora

Each of pronoun anaphora rules consists of the following main parts:

- list of pronouns that trigger the rule
- description of possible paths (via syntactic slots) from a possible controller to a pronoun
- description of possible properties of a controller
- a rule of agreement between a controller and a pronoun
- linear direction of the link (whether controller is to the left of the pronoun or to the right)
- value of the link

Мальчик любит этот дом — он его строил

Syntactic tree without non-tree links



As the pronouns (он, его) are found in the text, the anaphora rules are triggered and give such possible links as a result:

- link 1:

```
Proform "он"; ProformParent "строить";
ProformSlot Object_Direct; Controller "дом"
```

- link 2:

```
Proform "он"; ProformParent "строить";
ProformSlot Object_Direct; Controller "мальчик"
```

- link 3:

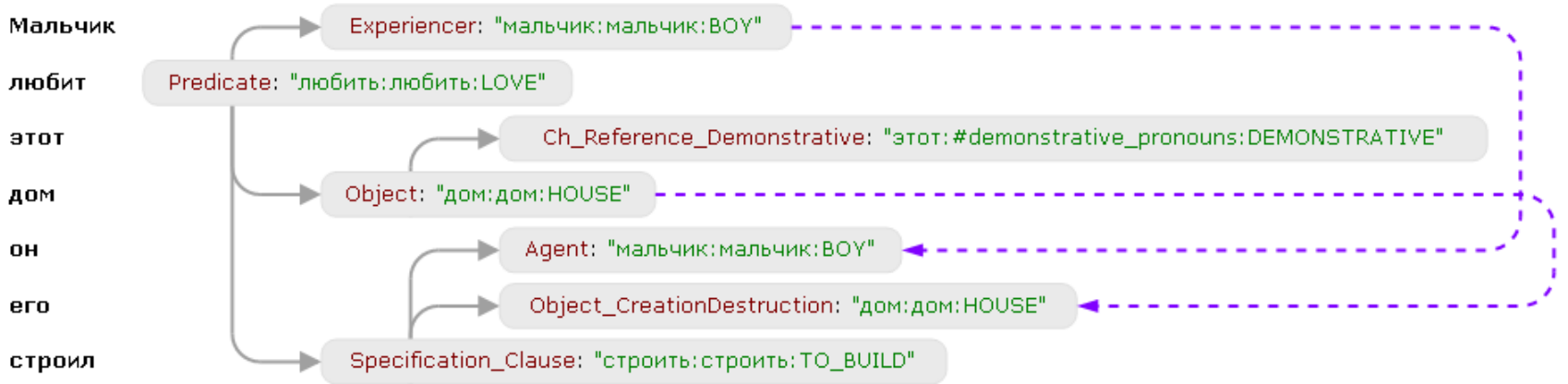
```
Proform "он"; ProformParent "строить";
ProformSlot Subject; Controller "мальчик"
```

- link 4:

```
Proform "он"; ProformParent "строить";
ProformSlot Subject; Controller "дом"
```


Мальчик любит этот дом — он его строил

Semantic tree with non-tree links



Relative anaphora

- Links of this type are also presented by special rules which have almost the same main parts as pronoun anaphora rules except that relative pronouns unlike personal pronouns must be controlled obligatorily
- If for a given relative pronoun a controller is not found then the structure is considered invalid.

Мальчик видит игрушку девочки, которая пришла

Semantic tree with non-tree links



Coreference

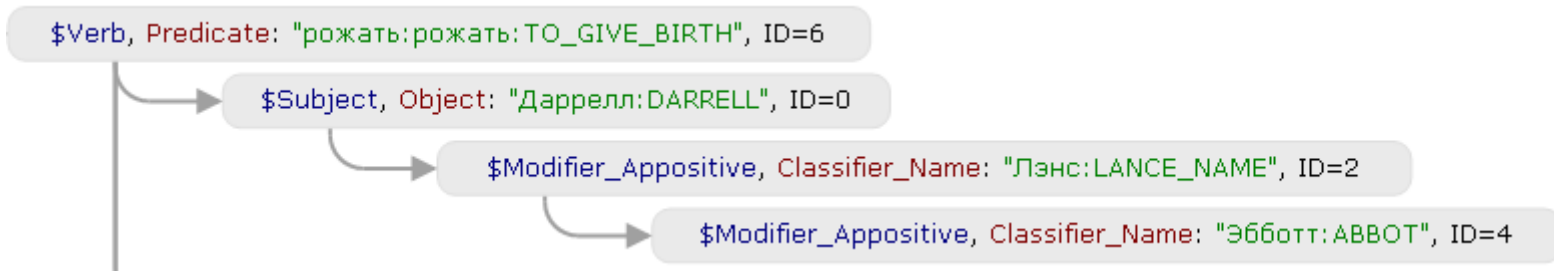
A two-stage process of coreference resolution:

- **Candidate extraction (recall-oriented phase)**
 - Named entity extraction (Person, Organization, Location)
 - Other potential coreferents extraction
- **Identification and filtering (precision-oriented phase)**
 - Candidates identification
 - Removal of the objects left without pairs

Named entity extraction

- Fully known proper names

Даррелл Лэнс Эбботт родился в Арлингтоне (штат Техас), пригороде Далласа и Форт Уорта

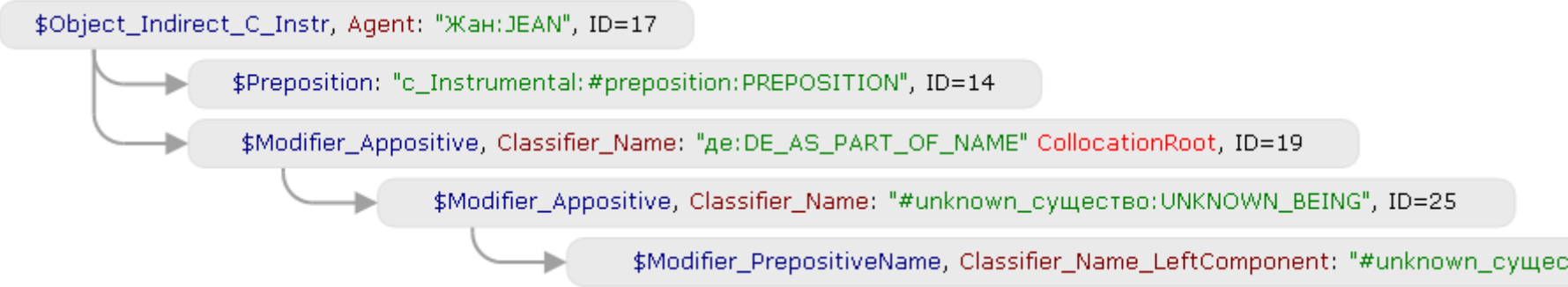


First name	Middle name	Surname
Даррелл	Лэнс	Эбботт

Named entity extraction

- Partially known proper names

Людовик XIV прислал корпус из 6 тысяч человек во главе с Жаном де Колиньи-Салиньи.

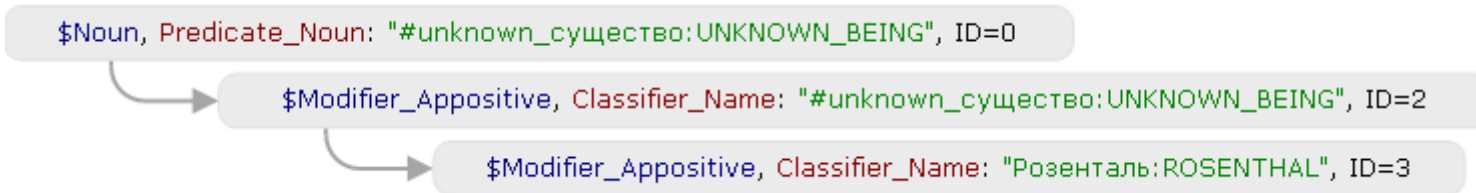


First name	Middle name	Surname
Жан		де Колиньи-Салиньи

Named entity extraction

- Partially known proper names

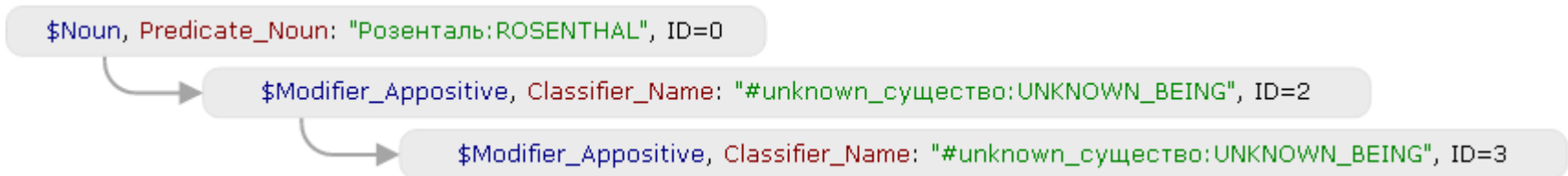
Дитмар Эльяшевич Розенталь



First name	Middle name	Surname
Дитмар	Эльяшевич	Розенталь

- Works the other way round (different tree, same attributes):

Розенталь Дитмар Эльяшевич



Named entity extraction

- Completely unknown proper names

HBO announced Wednesday morning that <...>

\$Verb, Predicate: "announce:announce:TO_ANNOUNCE", ID=2

\$Subject, Agent: "#Acronym:ACRONYM_", ID=0

<...> уничтожить отставший от основных сил корпус генерала Юэлла.

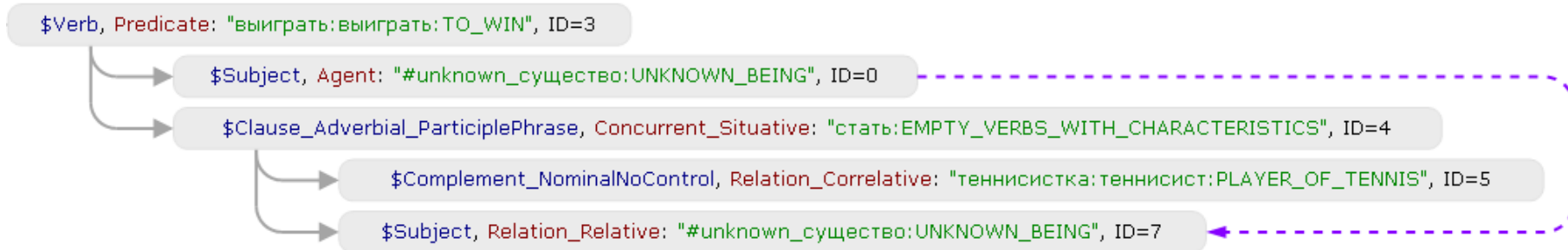
\$Object_Indirect_Genitive, Object_Relation: "генерал:генерал:GENERAL_MILITARY", ID=26

\$Modifier_Appositive, Classifier_Name: "#unknown_существо:UNKNOWN_BEING", ID=27

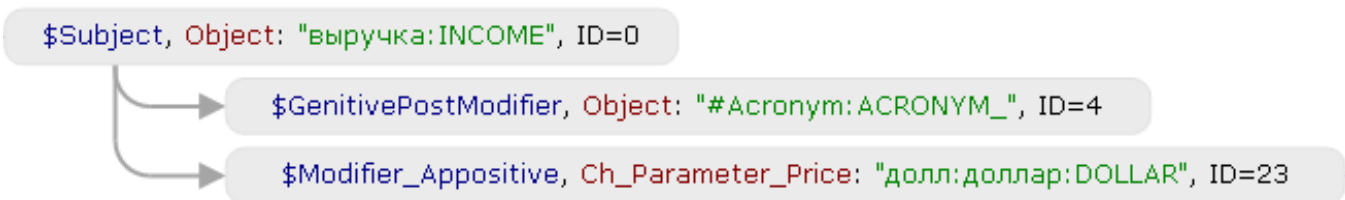
Named entity extraction

- Completely unknown proper names

<...> в Нэшвилле Месхи выигрывает свой первый турнир <...>, став первой советской теннисисткой <...>.



Выручка СВВ в 2005 году — 7,205 млрд долл.



Identification

Даррелл Лэнс Эбботт <..> Эбботта <..>

Surnames match, no obvious contradictions in other attributes

Даррел Лэнс Эбботт <..> Даррелом <..>

First names match, no obvious contradictions in other attributes

Даррел Лэнс Эбботт родился <..> в семье музыканта и продюсера Джерри Эбботта.

Surnames match, yet first names differ => different Person entities

Иванов получил зарплату. Иванова рада

Different gender => different entities

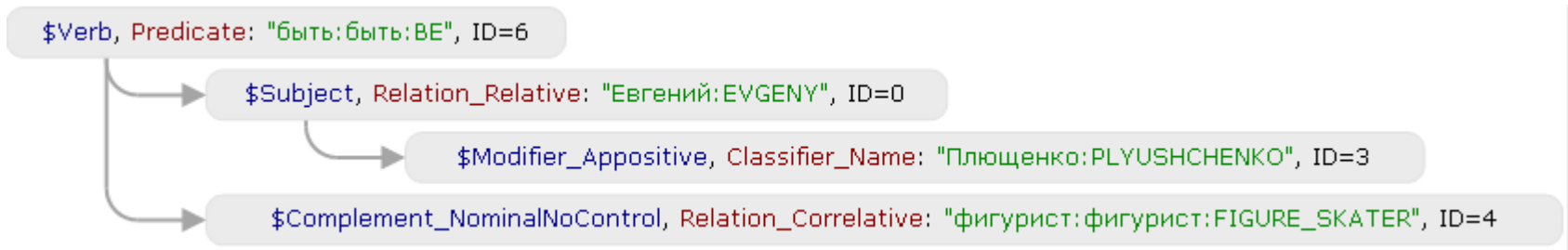
Иванов получил зарплату. Иванова обуяла радость

Surnames match, no obvious contradictions in other attributes

Identification: a pinch of hierarchy-based similarity

*Евгений Плющенко – единственный в мире **фигурист**, который смог завоевать медали четырёх зимних Олимпиад. **Свой** первый олимпийский опыт **спортсмен** получил в 2002 году на играх в американском Солт-Лейк-Сити.*

- Step 1 (tree structure):



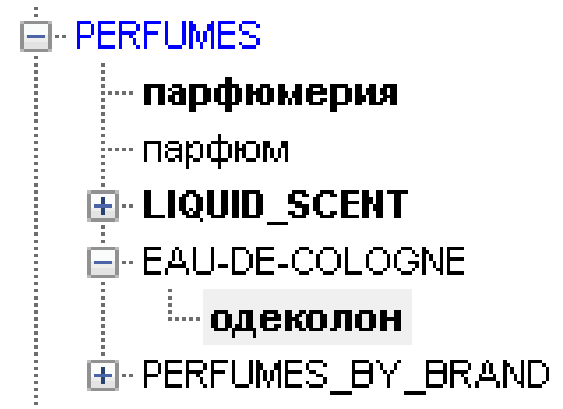
Identification: a pinch of hierarchy-based similarity

- Step 2 (hierarchy structure):



Non-proper names: more hierarchy-based similarity

*<...>каким-то дорогим **одеколоном**, **который** она приобрела для молодого супруга. Мы попросили понюхать **этот парфюм**.*



*<...>Скоро ужасную **клячу**, словно сбежавшую с живодерни увидели и другие зрители. Люди смеялись, удивлялись, спрашивали, негодовали. Как могла попасть сюда **эта лошадь**?*



Thanks for attention
Questions and comments are most
welcome