**.arff**

*poznámky sa označujú znakom:*

- aj prázdne riady v rámci nejakého slovného popisu musia začínať týmto znakom

*chýbajúce hodnoty sa označujú znakom:* ?

*v samotnej dátovej časti sú jednotlivé atribúty oddelené:*

**Náležitosti (povinné časti) arff dokumentu:**

1.

2.

3.

**Typy premenných:**

1. numeric (NUMERIC)

a) real – hodnoty, ktoré môžu byť aj desatinné

b) integer – celočíselné hodnoty

2. {nominal} – ide o opakujúcu sa hodnotu znaku, napr. pohlavie, vzdelanie, nejaká kategória, ...

– obmeny premennej sú jednoznačne stanovené v kučeravých zátvorkách

– softvér je citlivý na veľké a malé písmená

– ak ide o *viacslovnú* obmenu znaku, je potrebné daný výraz dať do *úvodzoviek,* napr. ‘Banska Bystrica’ alebo použiť pomlčku, či podčiarkovník

3. string (STRING) – textová premenná

4. date (DATE) – presný formát dátumu musí byť pri názve atribútu definovaný v úvodzovkách a musí mať rovnaký tvar ako samotné dáta

*napr.* máme údaje o predaji výrobku a tvar záznamov je "yyyy-mm-dd hh:mm:ss" pričom presne takýto tvar musia mať ak konkrétne záznamy, teda "2016-08-31 12:03:42"

**Vytvorený arff dokument vyzerá teda nasledovne:**

@relation <relation-name>

@attribute <attribute-name> <datatype>

@data