

Finding inconsistent and invalid data

This is the collapsible text.

For various reasons, such as data import or bugs of some version of the interface, a CASTEMO knowledge graph can contain **inconsistent data**. It is thus important to identify this data and correct the inconsistencies either manually or from a script.

Get all entities involved in more than one synonym cloud

In the CASTEMO data model, one entity can be only involved in one synonym cloud.

ReQL

```
r.db("inkvisitor")
  .table("relations")
  .filter({ type: "SYN" })
  .getField("entityIds")
  .reduce(function (acc, val) {
    return acc.setUnion(val);
  })
  .default([])
  .map(function (ide) {
    return {
      id: ide,
      count: r
        .db("inkvisitor")
        .table("relations")
        .filter({ type: "SYN" })
        .getField("entityIds")
        .filter(function (syn) {
          return syn.contains(ide);
        })
        .count()
    };
  })
```

```
    })  
    .count(),  
  };  
})  
.filter(function (a) {  
  return a("count").gt(1);  
});
```

R

Empty language: get labels of all concepts with empty language

ReQL

```
r.db("inkvisitor")  
  .table("entities")  
  .filter({ language: "", class: "C" })  
  .getField("label");
```

Revision #5

Created 20 March 2023 11:40:57 by Prof. Dr. David ZbÍral, Ph.D.

Updated 21 March 2023 13:53:21 by Gideon Kotzé PhD