

Appendix: BIB-R: a Benchmark for the Interpretation of Bibliographic Records

Joffrey Decourselle¹, Fabien Duchateau¹, Trond Aalberg², Naimdjon Takhirov³, and Nicolas Lumineau¹

¹ LIRIS, UMR5205, Université Claude Bernard Lyon 1, Lyon, France
`firstname.lastname@liris.cnrs.fr`

² NTNU, Trondheim, Norway `tronndaal@idi.ntnu.no`

³ Westerdals - Oslo School of Arts, Communication and Technology - Faculty of Technology, Oslo, Norway `taknai@westerdals.no`

Abstract. This appendix contains all experiment results for the submission to TPDL2016. In the cultural heritage domain, cataloging items is a crucial task, which has been performed in decades based on the MAchine Readable Cataloguing (MARC) format. In a context which promotes the use of semantics and sharing of information, MARC has demonstrated its limitations. The Functional Requirements for Bibliographic Records (FRBR), expected to be a successor of MARC, is a semantic model for representing cultural items. The complex transformation of MARC catalogs to FRBR catalogs (FRBRization) led to the proposition of various tools and approaches. However these projects and the results they achieve are difficult to compare on a fair basis due to a lack of common datasets and appropriate metrics. Our contributions fill this gap by proposing the first public benchmark for the FRBRization process.

Keywords: benchmark, migration, record interpretation, FRBRization, LRM, FRBR, MARC, dataset, evaluation metric

1 Formal notation of pre-FRBRization metrics

Table 1 provides formal notations for the predictive metrics. Each metric computes a percentage defined as the number of records concerned with the given pattern/issue divided by the total number of records. The detection of these issues uses the first-order logic notation for genericity reasons. In the table, we define a record r which belongs to a set of records \mathcal{R} . We note $r \hookrightarrow$ 'specification' the fact that the record r implements a given specification such as catalog rule (e.g., AACR2) or punctuation (e.g., ISBD). This record is composed of a set of fields, each of them representing a concept. For instance, the notation $f \in r, f \rightsquigarrow \text{'title'}$ indicates that the field f belongs to the record r and it stands for a *title* in this context. We also introduce a rule ℓ from a set \mathcal{L} . A rule contains a condition denoted $cond_\ell$, which satisfies a field f when $cond_\ell \models f$. Some fields may encode specific values that needs to be extracted. We define $\varphi(value)$ as a function which extracts frequent patterns from a given *value* by analyzing the whole collection. It returns an empty set if no pattern is detected, or a list of pattern otherwise (e.g., "illustrated by"). Various approaches enable this extraction of patterns, traditionally based on machine learning techniques [2]. To illustrate the table, let us see the **metric MID**, which stands for the percentage of records in which the record identifier is missing. Its formal notation means that whatever the field f in a record r , this field f does not correspond to the concept 'record identifier'. Since this is a generic definition, it can be instantiated for a specific format. For example, the UNIMARC notation for the MID metric could be written: $\forall f \in r, f \neq '001'$. In MARC21, the **metric MPD** about missing publication date can be redefined as: $\forall f \in r, f \neq '260c' . A more complex notation deals with the **metric CPN**, which consists of measuring the percentage of records that implements a specification (catalog rule or punctuation), or which contains a field f used to store local data (traditionally the 900 fields in MARC21) or a field with recurrent values (e.g., the Norwegian

country code could be written "NO" or "NOR"). Finally, bibliographic patterns can be detected using multiple fields. Note that when we indicate $value_f \in ['Derivations']$, it means that the value of f corresponds to a Derivation (e.g., trl for translator in MARC21⁴) according to Tillet and Riva's taxonomies [1]. The metric MR can be decomposed into more detailed metrics such as **MR-AUG** to calculate the percentage of missing rules for detecting all augmentations. In the same fashion, we define the metrics **MR-DER** for derivations, **MR-AGG** for aggregations, **MR-COW** for complementary works, and **MR-CPN** for cataloging practices. Their formal notation is obtained by combining the definitions of both MR and the given pattern/issue.

2 Formal notation of post-FRBRization metrics

Table 2 provides first-order logic notations for the detection of post-FRBRization issues. During evaluation, we compare two collections, \mathcal{T} which is produced by a tool and the expert collection \mathcal{E} . This comparison depends on the type of data. Consider the data $e \in \mathcal{E}$ and $t \in \mathcal{T}$. When dealing with entities, the type of entity and the value of its main label (e.g., title, name) needs to be verified:

$$e \equiv t \iff type_e = type_t \wedge value_e = value_t$$

For relationships, the checking is performed based on the type of relationship and the two linked entities:

$$e \equiv t \iff type_e = type_t \wedge entity_e^1 = entity_t^1 \wedge entity_e^2 = entity_t^2$$

Finally, the properties are compared according to their type, their owner (entity) and their value:

$$e \equiv t \iff type_e = type_t \wedge entity_e = entity_t \wedge value_e = value_t$$

Besides, it is not possible to verify information about bibliographic patterns without annotation in the expert collection. Thus we define the set $\mathcal{E}' \in \mathcal{E}$ which includes all main elements of a pattern (i.e., the main entity and the main relationship) and the set $\mathcal{E}'' \in \mathcal{E}$ which contains all secondary elements of a pattern.

Let us describe Table 2. First, the **metric MD** is related to the missing data issue. The formal notation states that such data e appears in the expert collection \mathcal{E} but has no equivalence in the tool's collection \mathcal{T} . This metric computes the ratio between the number of missing data and the total number of data in the expert collection. It can be redefined for each type of data, i.e., MD-E for entities, MD-R for relationships and MD-P for properties. The **metric IAD** deals with incorrectly added data, i.e., which appear in \mathcal{T} but not in \mathcal{E} . It is defined as the number of incorrect data in \mathcal{T} divided by the total number of data in \mathcal{T} . Similarly to MD, the metric IAD can be redefined according to the data type. The **metric DLE** relates to errors in external link, i.e., either the link does not exist in \mathcal{E} or it has a different value for the same external source. The metric calculates precision, i.e., the number of erroneous links in \mathcal{T} divided by the total number of links in \mathcal{T} . The **metric SMD** computes the amount of data with a different semantics (usually a subsumption noted $t \subset e$) with regards to the total amount of data. The **metric MEND** measures the percentage of main entities in \mathcal{T} that have been correctly detected among all main entities in \mathcal{E}' . To check the correctness of a detection, a main entity e' should have an equivalent entity t in the tool's collection. Note that the **metrics MRND and ESE** have a similar definition due to our generic notation.

⁴ MARC21 code list for relators

	Related pattern/issue	Formal notation
AUG	Augmentation	$f \in r, (f \rightsquigarrow \text{'secondary resp.'}) \vee (f \rightsquigarrow \text{'title'} \wedge \varphi(\text{value}_f) \neq \emptyset) \vee (f \rightsquigarrow \text{'relator code'}) \wedge \text{value}_f \in [\text{'Augmentations'}])$
DER	Derivation	$f \in r, (f \rightsquigarrow \text{'original language'}) \wedge \exists f_2 \in r \wedge (f \rightsquigarrow \text{'resource language'}) \vee (f \rightsquigarrow \text{'title'}) \wedge \varphi(\text{value}_f) \neq \emptyset \vee (f \rightsquigarrow \text{'relator code'}) \wedge \text{value}_f \in [\text{'Derivations'}]) \vee (f \rightsquigarrow \text{'variant title'})$
AGG	Aggregation	$f \in r, (f \rightsquigarrow \text{'collection/ensemble link'}) \vee (f \rightsquigarrow \text{'collective title'}) \vee (f \rightsquigarrow \text{'proper title'}) \wedge \exists f^1, \dots, f^k \in r \wedge f^1 \rightsquigarrow \text{'proper title'} \wedge \dots \wedge f^k \rightsquigarrow \text{'proper title'}) \vee (f \rightsquigarrow \text{'part title'}) \vee (f \rightsquigarrow \text{'variant title'})$
COW	Complementary works	$f \in r, (f \rightsquigarrow \text{'title'}) \wedge \exists f_2 (f_2 \rightsquigarrow \text{'subtitle'}) \wedge \text{value}_{f_2} \neq \text{value}_f) \vee (f \rightsquigarrow \text{'title'}) \wedge \varphi(\text{value}_f) \neq \emptyset \vee (f \rightsquigarrow \text{'linking field'}) \vee (f \rightsquigarrow \text{'note'}) \vee (f \rightsquigarrow \text{'added entry'})$
MID	Missing record identifier	$\forall f \in r, f \not\rightsquigarrow \text{'record identifier'}$
MPD	Missing publication date	$\forall f \in r, f \not\rightsquigarrow \text{'publication date'}$
MTF	Missing type and form	$\forall f \in r, f \not\rightsquigarrow \text{'type/form'}$
TLE	Title linkage error	$f \in r, f \rightsquigarrow \text{'auth. title'} \wedge \nexists r_2 \in \mathcal{R} (f_2 \in r_2 \wedge f_2 \rightsquigarrow \text{'record identifier'}) \wedge \text{value}_{f_2} = \text{value}_f)$
MUT	Missing uniform title	$\forall f \in r, f \not\rightsquigarrow \text{'uniform title'}$
MOT	Missing original title	$\forall f \in r, f \not\rightsquigarrow \text{'original title'}$
RLE	Resp. linking error	$f \in r, f \rightsquigarrow \text{'auth. resp.'} \wedge \nexists r_2 \in \mathcal{R} (f_2 \in r_2 \wedge f_2 \rightsquigarrow \text{'record identifier'}) \wedge \text{value}_{f_2} = \text{value}_f)$
MRC	Missing relator code	$\forall f \in r, f \not\rightsquigarrow \text{'relator code'}$
MAR	Missing auth. resp.	$\forall f \in r, f \not\rightsquigarrow \text{'auth. resp.'}$
CPN	Cataloguing practices and norms	$r \rightsquigarrow \text{'catalog rule'} \vee r \rightsquigarrow \text{'punctuation'} \vee (f \in r, f \rightsquigarrow \text{'local data'}) \vee \varphi(\text{value}_f) \neq \emptyset$
MR	Missing rule	$f \in r, \forall \ell \in \mathcal{L}, \text{cond}_\ell \not\models f$
UR	Useless rule	$\ell \in \mathcal{L}, \forall r \in \mathcal{R}, \forall f \in r, \text{cond}_\ell \not\models f$
CR	Conflicting rules	$\forall \ell_1 \in \mathcal{L}, \forall \ell_2 \in \mathcal{L}, \ell_1 \neq \ell_2 \wedge \text{cond}_{\ell_1} \equiv \text{cond}_{\ell_2}$

Table 1. List of pre-FRBRization metrics

	Related issue	Formal notation
MD	Missing data	$e \in \mathcal{E}, \forall t \in \mathcal{T}, t \not\equiv e$
IAD	Incorrectly added data	$t \in \mathcal{T}, \forall e \in \mathcal{E}, e \not\equiv t$
DLE	Data linking error	$t \in \mathcal{T}, t \rightsquigarrow \text{'external link'} \wedge (\exists e \in \mathcal{E} \wedge e \rightsquigarrow \text{'external link'} \wedge value_t \neq value_e \wedge source_t = source_e)$
SMD	Semantic mismatch data	$e \in \mathcal{E}, t \in \mathcal{T}, (t \subset e) \vee (e \subset t)$
MEND	Main entity not detected	$e' \in \mathcal{E}', \forall t \in \mathcal{T}, t \not\equiv e'$
MRND	Main relationship not detected	$e' \in \mathcal{E}', \forall t \in \mathcal{T}, t \not\equiv e'$
ESE	Error(s) in secondary elements	$e'' \in \mathcal{E}'', \forall t \in \mathcal{T}, t \not\equiv e''$

Table 2. List of post-FRBRization metrics

3 Assessing strengths and weaknesses - results by test

This first experiment aims at demonstrating the benefit of the dataset T42 when it comes to evaluating the strengths and weaknesses of FRBRization tools. For the three tools, we have run each test from the dataset T42 and the evaluation is performed using post-FRBRization metrics. In other words, each tool has produced a FRBR collection for each test, and these generated FRBR collections have been compared to the expert ones provided in the benchmark. A basic set of rules is available with each tool. For equity reasons, we have not tuned the tools by updating their set of rules.

In the following plots, the results are organized by tool and by test.

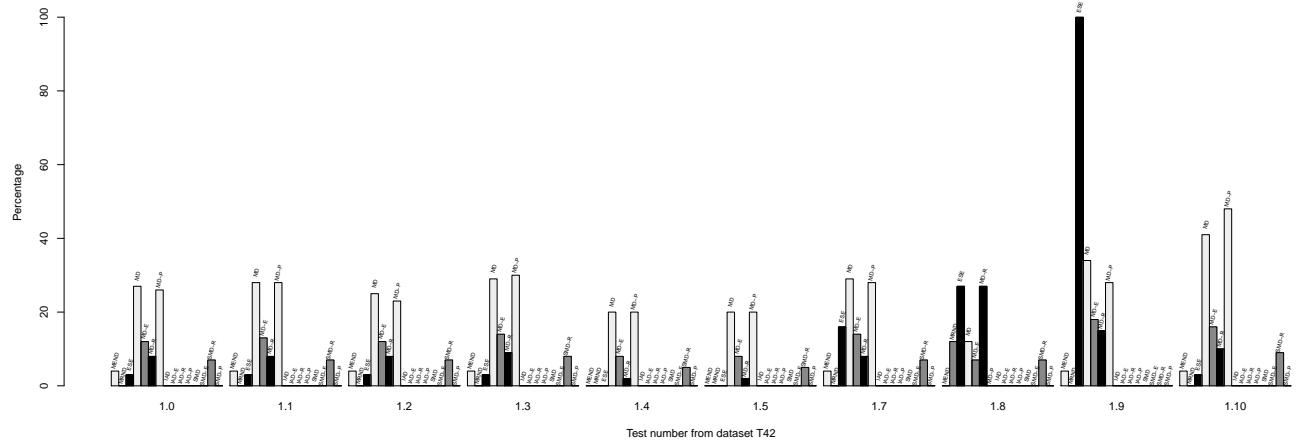


Fig. 1. Quality results for FRBR-ML with dataset T42 tests 1.x

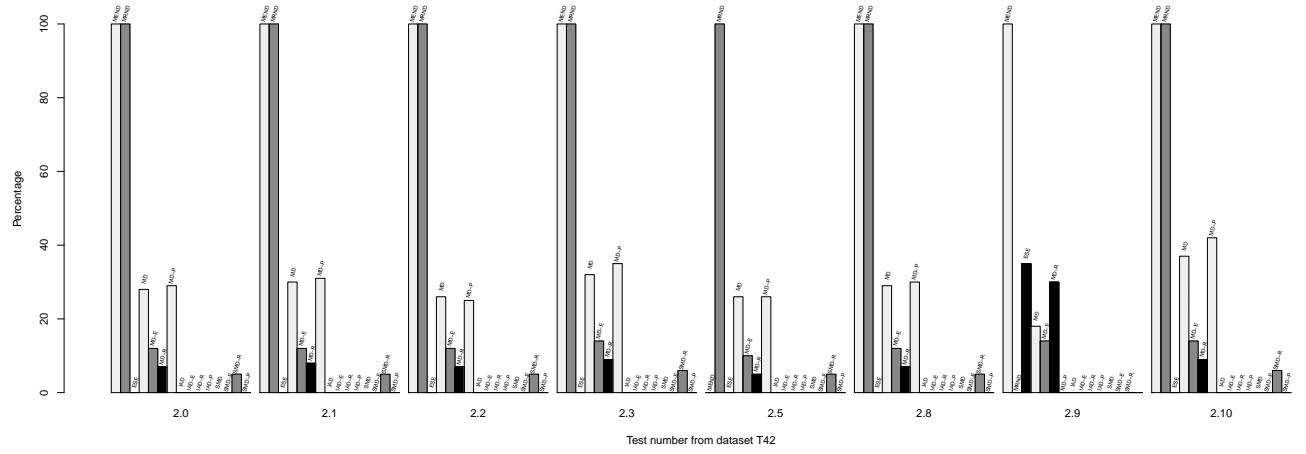


Fig. 2. Quality results for FRBR-ML with dataset T42 tests 2.x

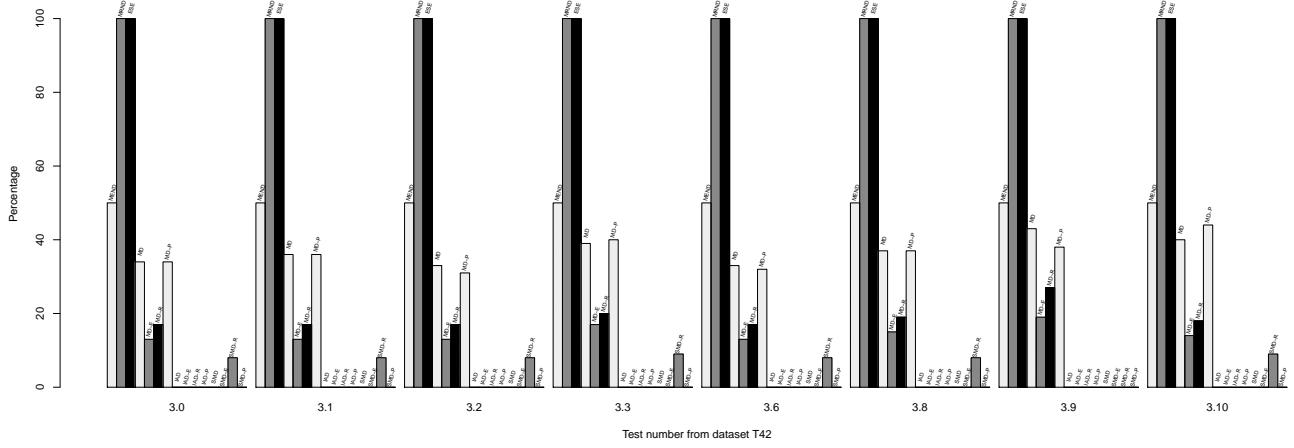


Fig. 3. Quality results for FRBR-ML with dataset T42 tests 3.x

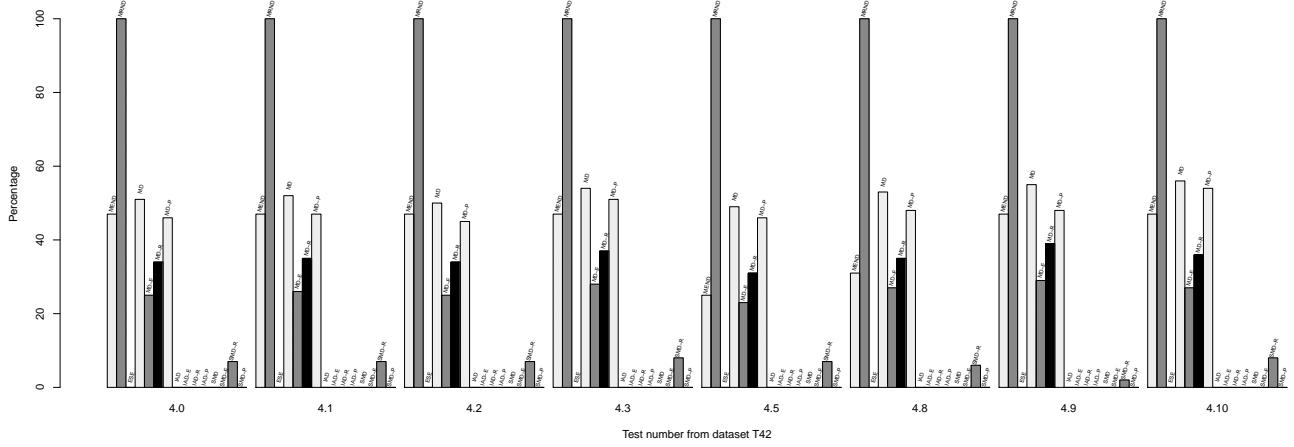


Fig. 4. Quality results for FRBR-ML with dataset T42 tests 4.x

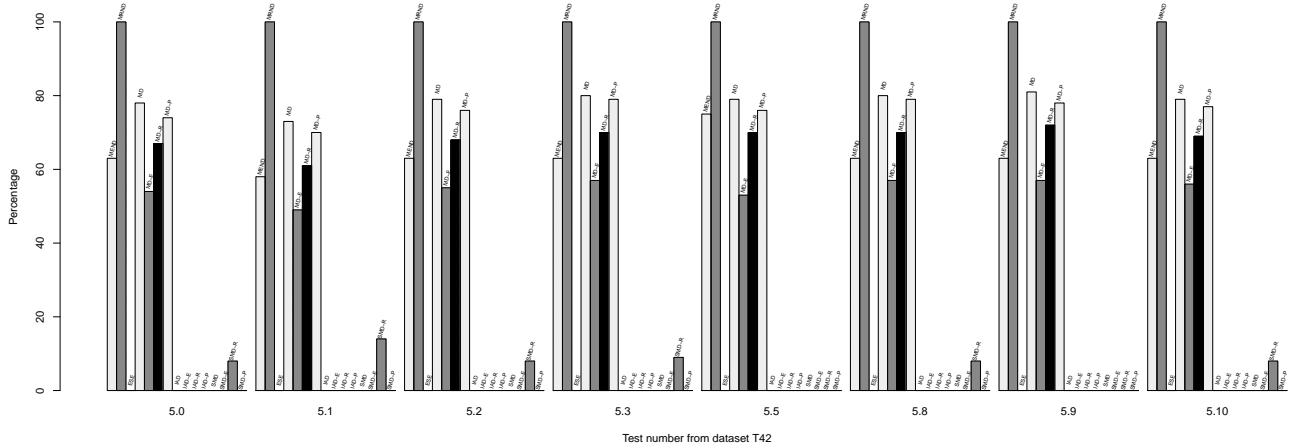


Fig. 5. Quality results for FRBR-ML with dataset T42 tests 5.x

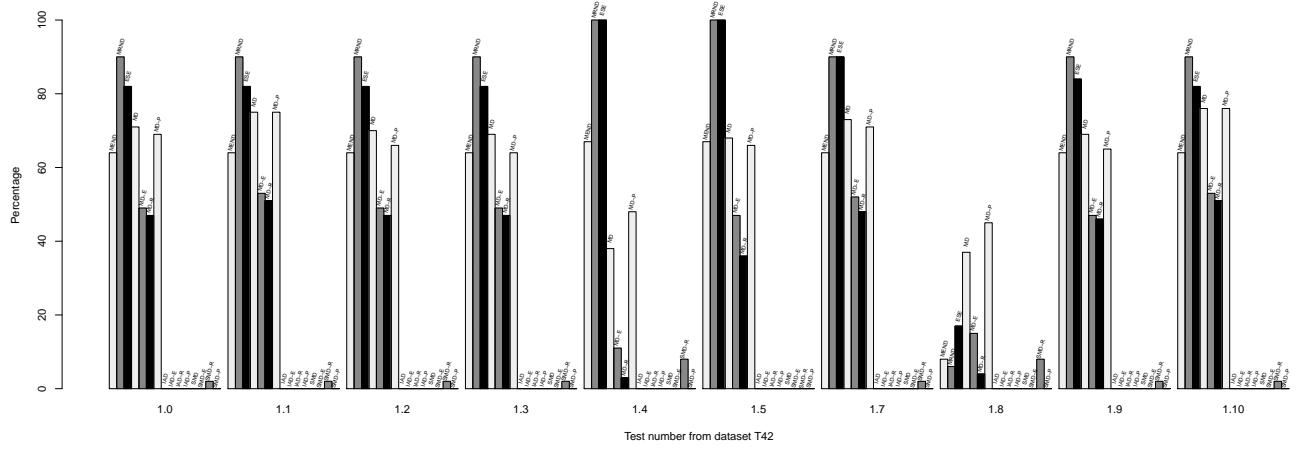


Fig. 6. Quality results for Variation/VFRBR with dataset T42 tests 1.x

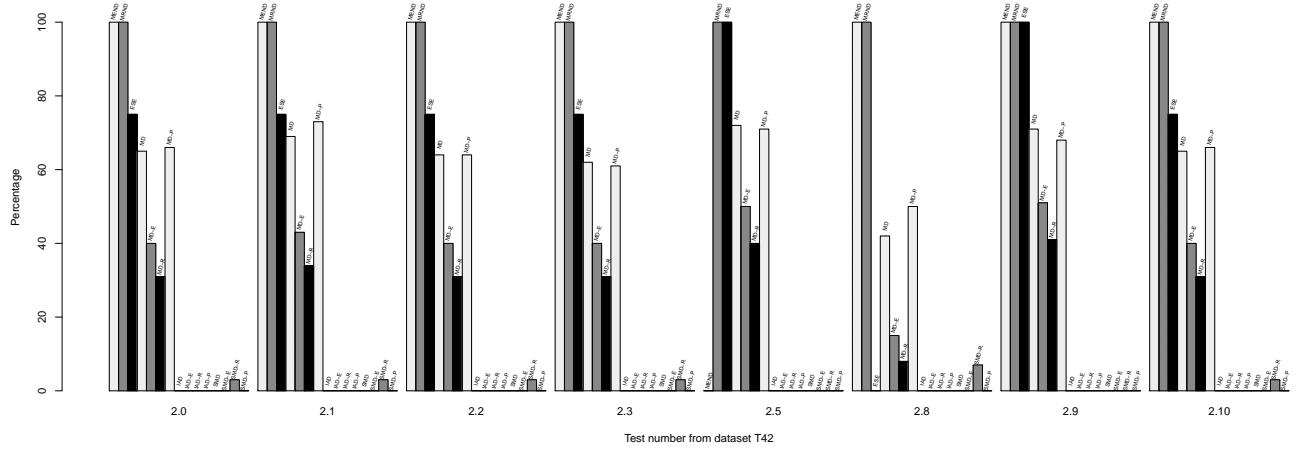


Fig. 7. Quality results for Variation/VFRBR with dataset T42 tests 2.x

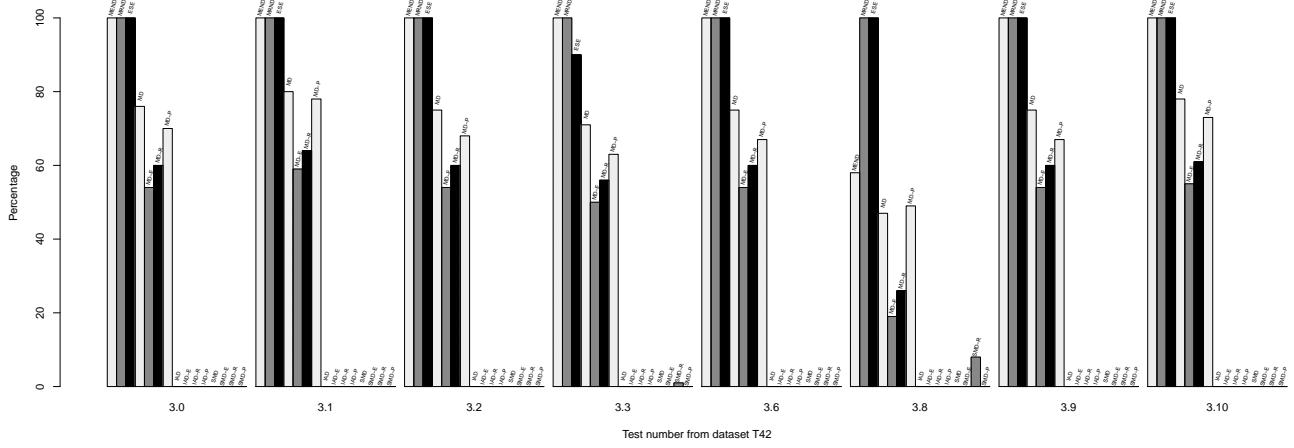


Fig. 8. Quality results for Variation/VFRBR with dataset T42 tests 3.x

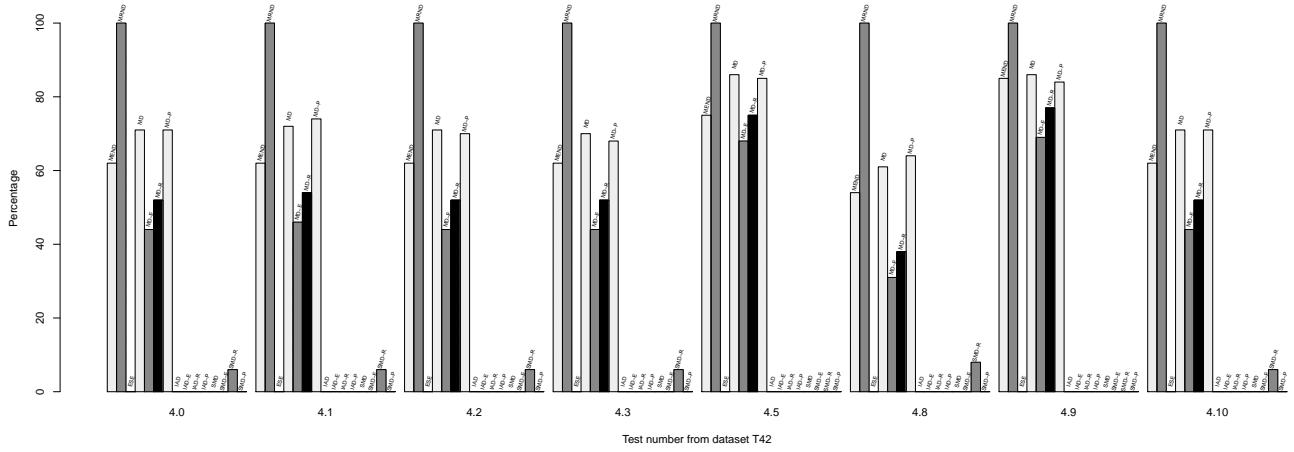


Fig. 9. Quality results for Variation/VFRBR with dataset T42 tests 4.x

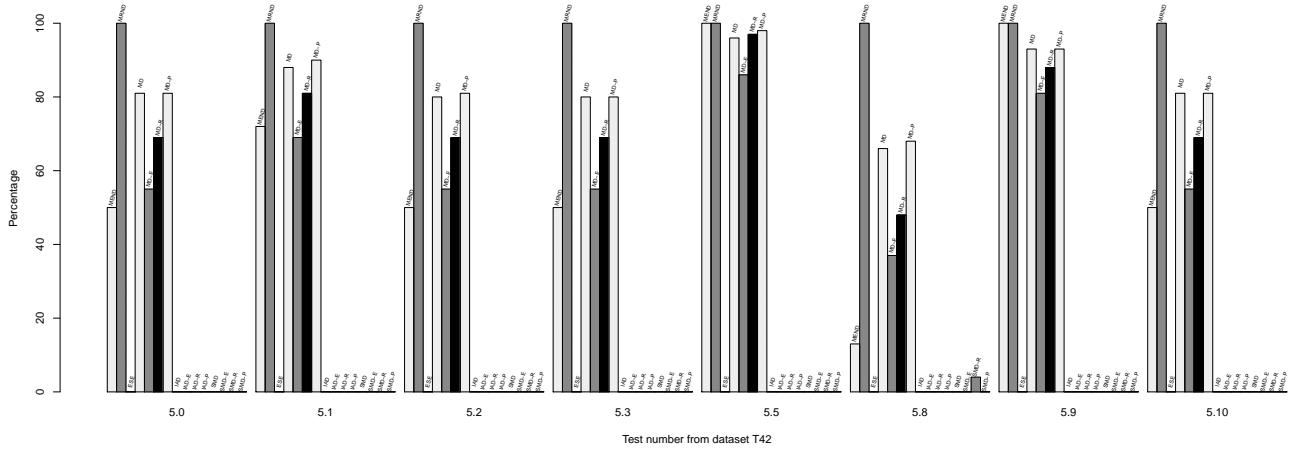


Fig. 10. Quality results for Variation/VFRBR with dataset T42 tests 5.x

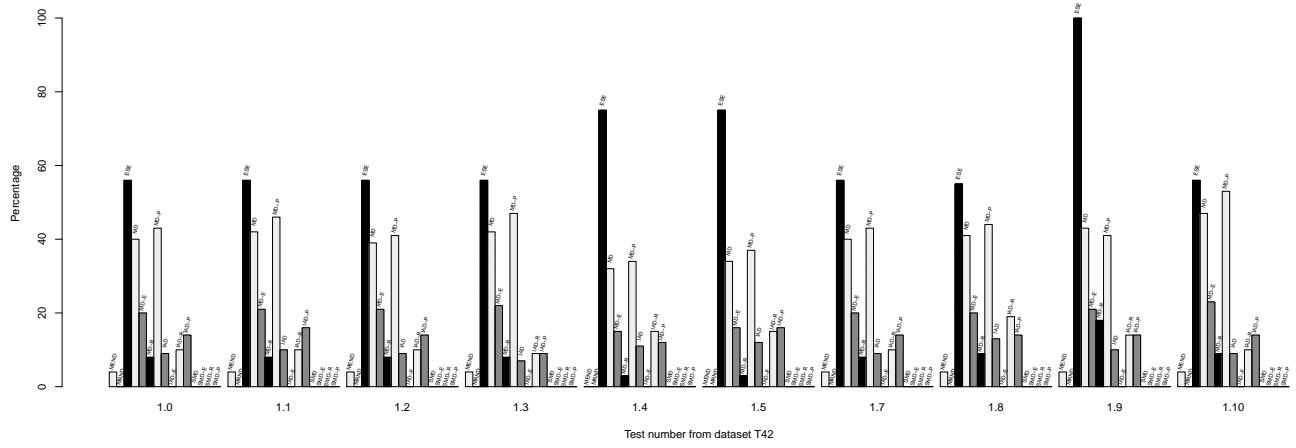


Fig. 11. Quality results for Extensible Catalog with dataset T42 tests 1.x

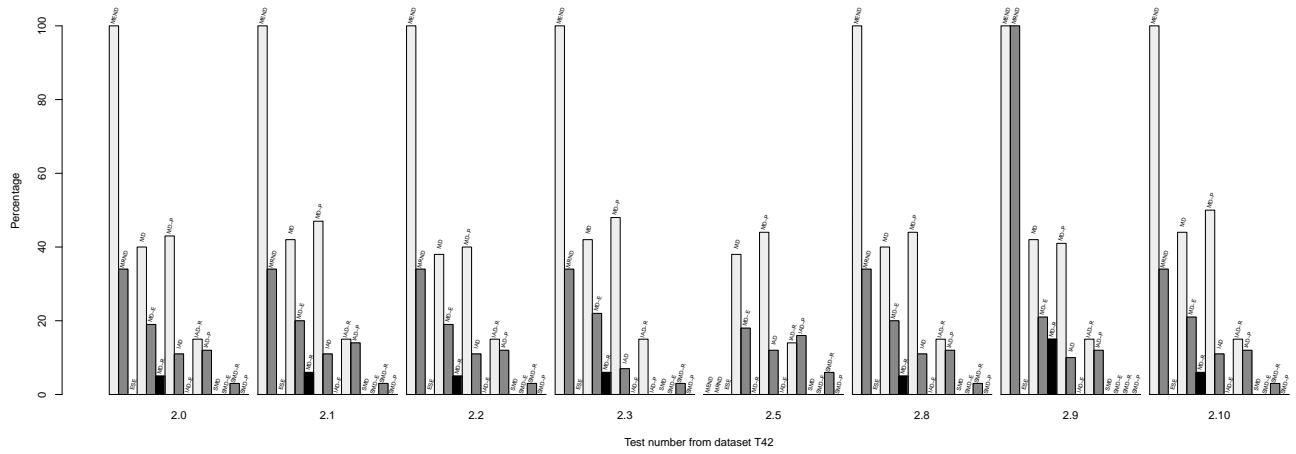


Fig. 12. Quality results for Extensible Catalog with dataset T42 tests 2.x

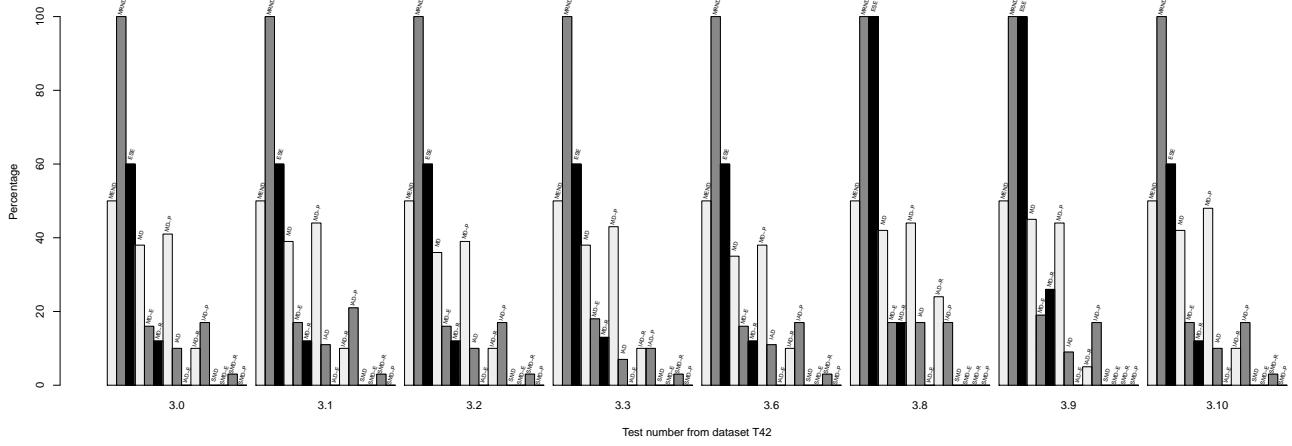


Fig. 13. Quality results for Extensible Catalog with dataset T42 tests 3.x

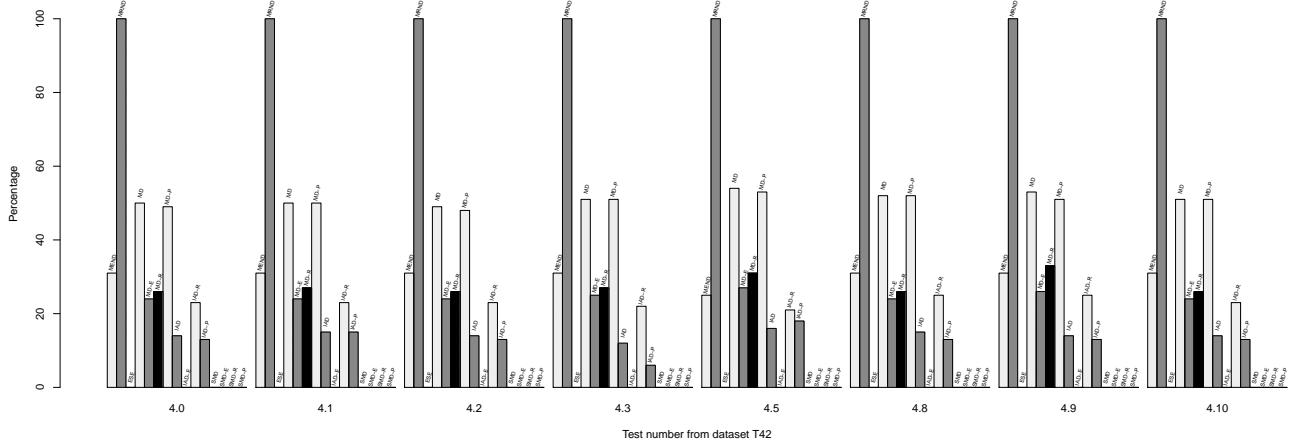


Fig. 14. Quality results for Extensible Catalog with dataset T42 tests 4.x

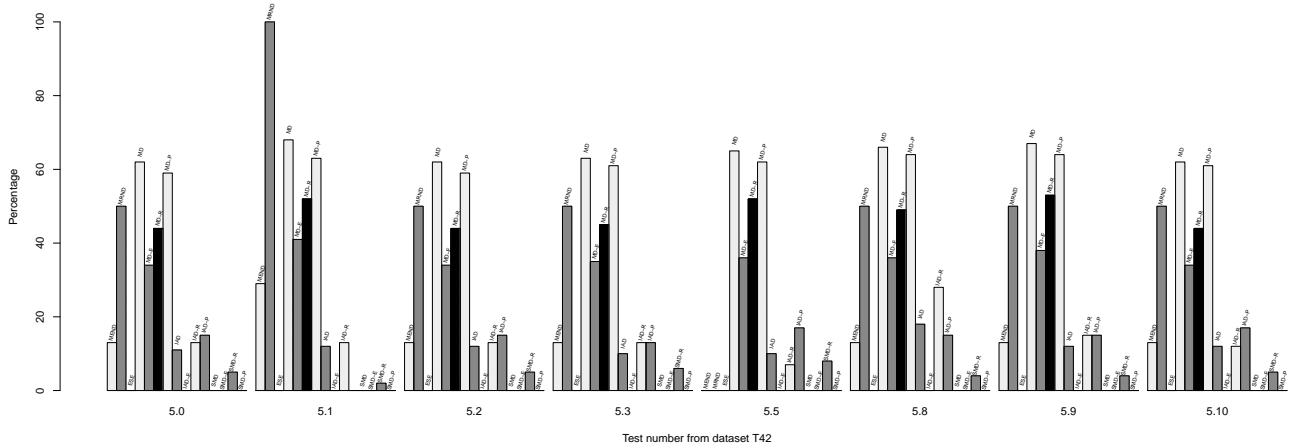


Fig. 15. Quality results for Extensible Catalog with dataset T42 tests 5.x

4 Assessing strengths and weaknesses - results by metric

This first experiment aims at demonstrating the benefit of the dataset T42 when it comes to evaluating the strengths and weaknesses of FRBRization tools. For the three tools, we have run each test from the dataset T42 and the evaluation is performed using post-FRBRization metrics. In other words, each tool has produced a FRBR collection for each test, and these generated FRBR collections have been compared to the expert ones provided in the benchmark. A basic set of rules is available with each tool. For equity reasons, we have not tuned the tools by updating their set of rules.

In the following plots, the results are organized by tool and by metric.

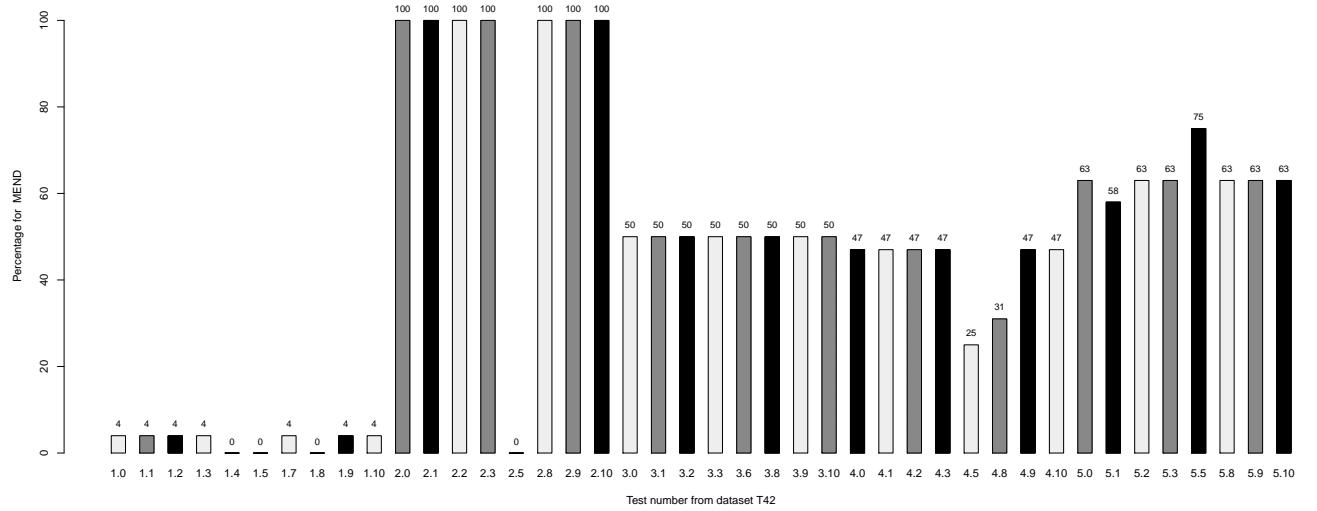


Fig. 16. Quality results for FRBR-ML with dataset T42 and metric MEND

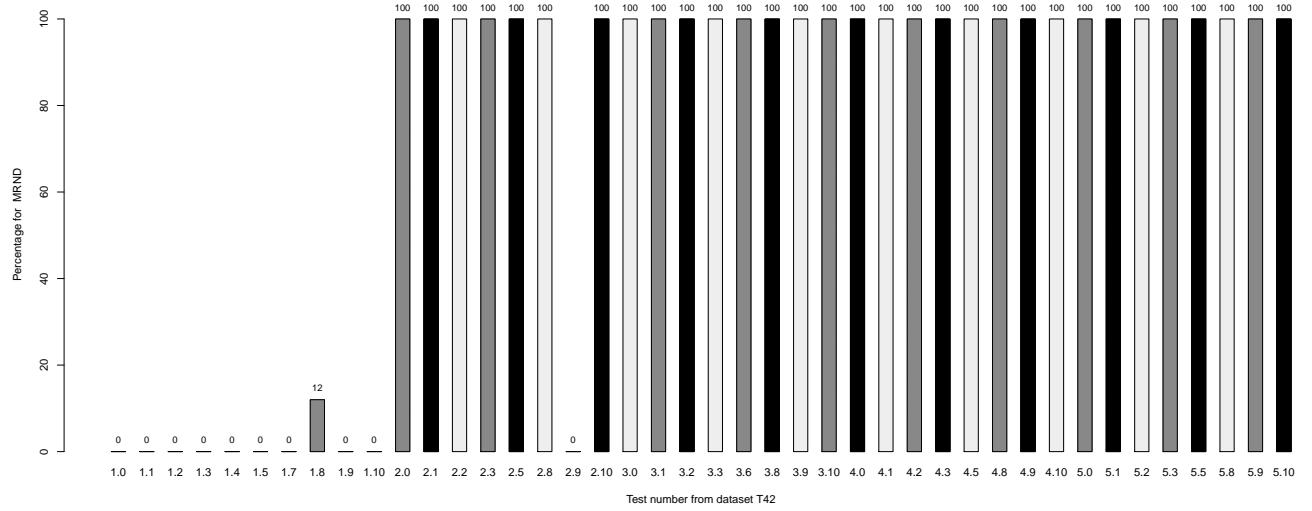


Fig. 17. Quality results for FRBR-ML with dataset T42 and metric MRND

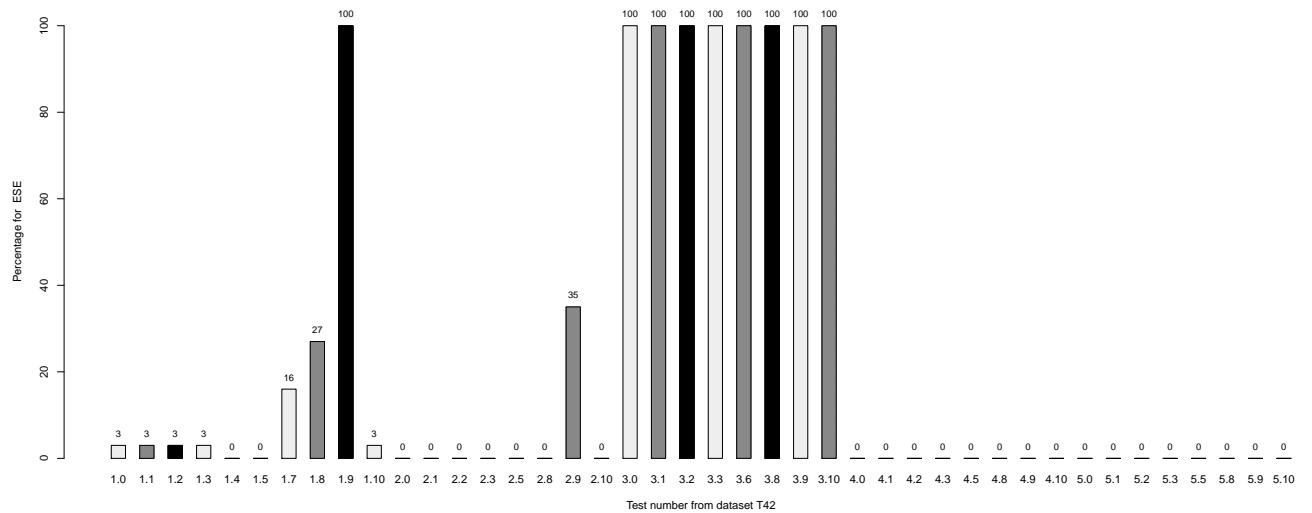


Fig. 18. Quality results for FRBR-ML with dataset T42 and metric ESE

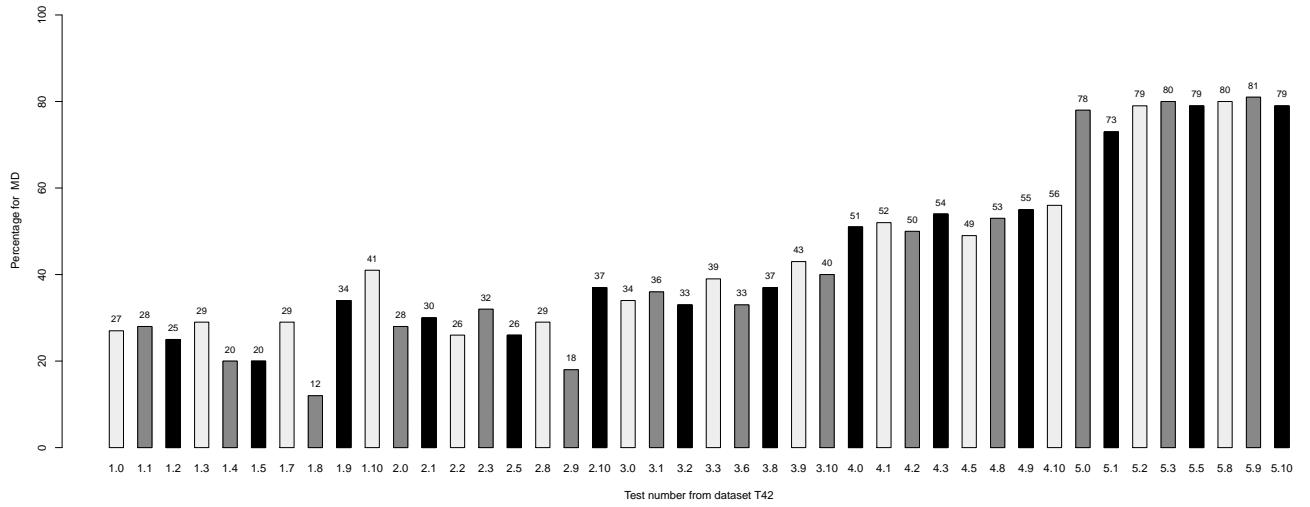


Fig. 19. Quality results for FRBR-ML with dataset T42 and metric MD

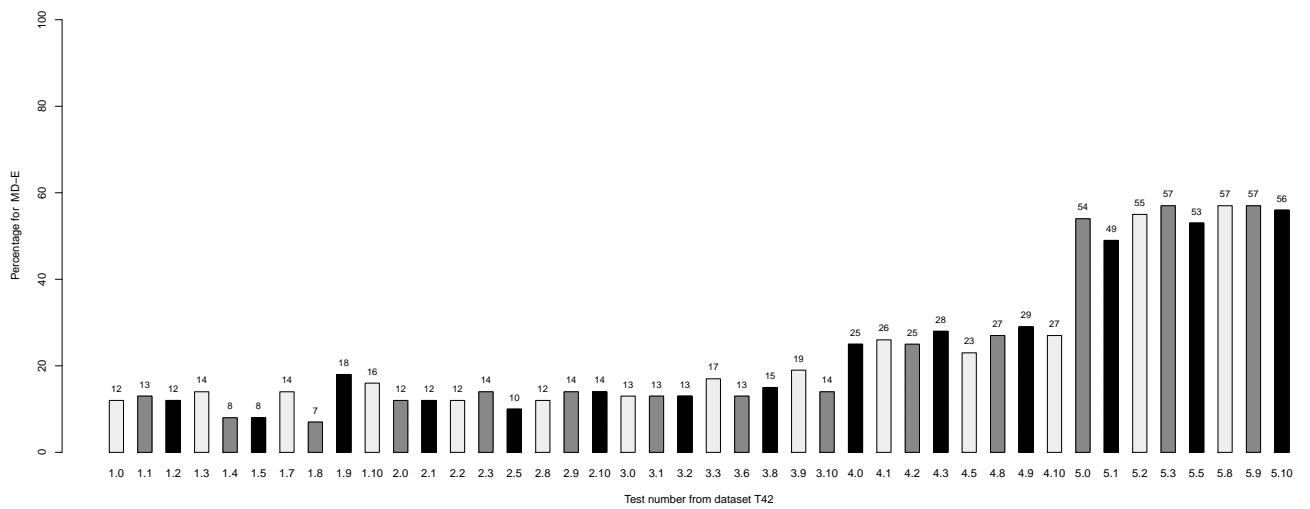


Fig. 20. Quality results for FRBR-ML with dataset T42 and metric MD-E

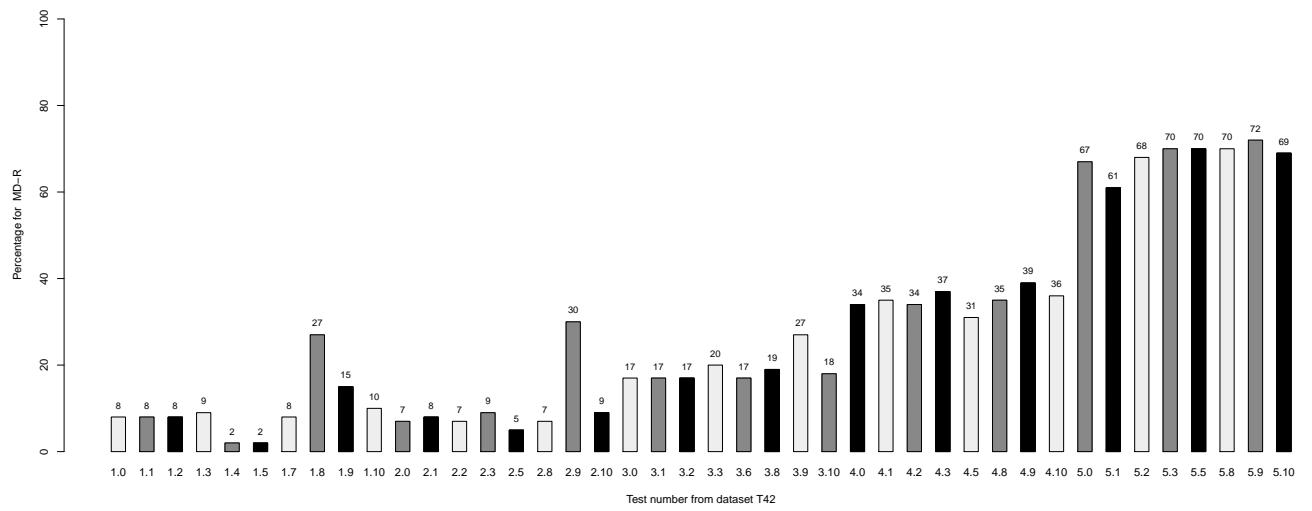


Fig. 21. Quality results for FRBR-ML with dataset T42 and metric MD-R

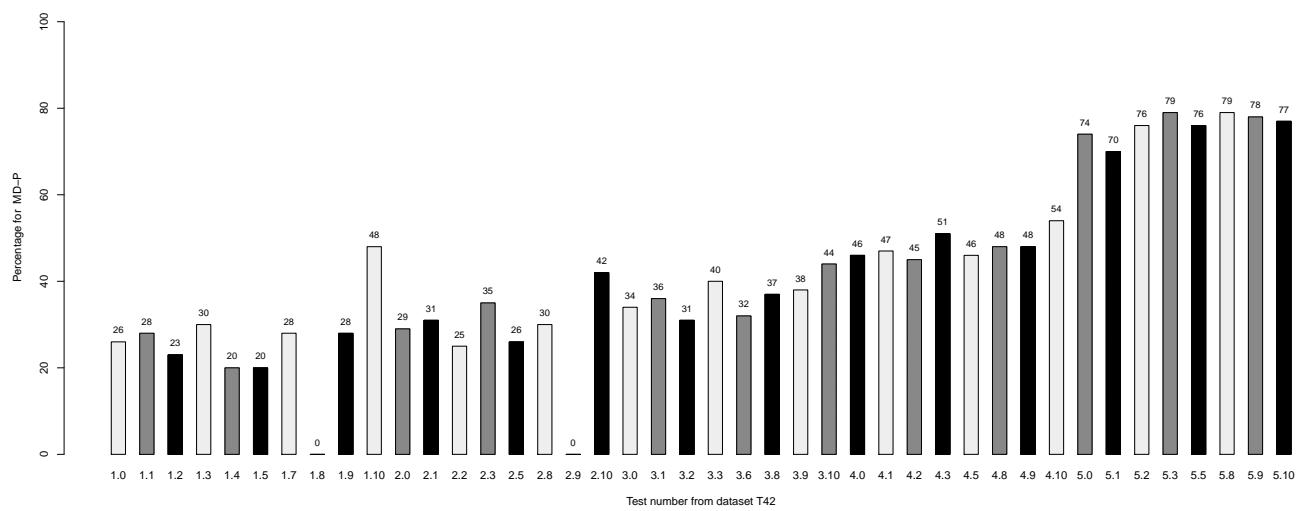


Fig. 22. Quality results for FRBR-ML with dataset T42 and metric MD-P

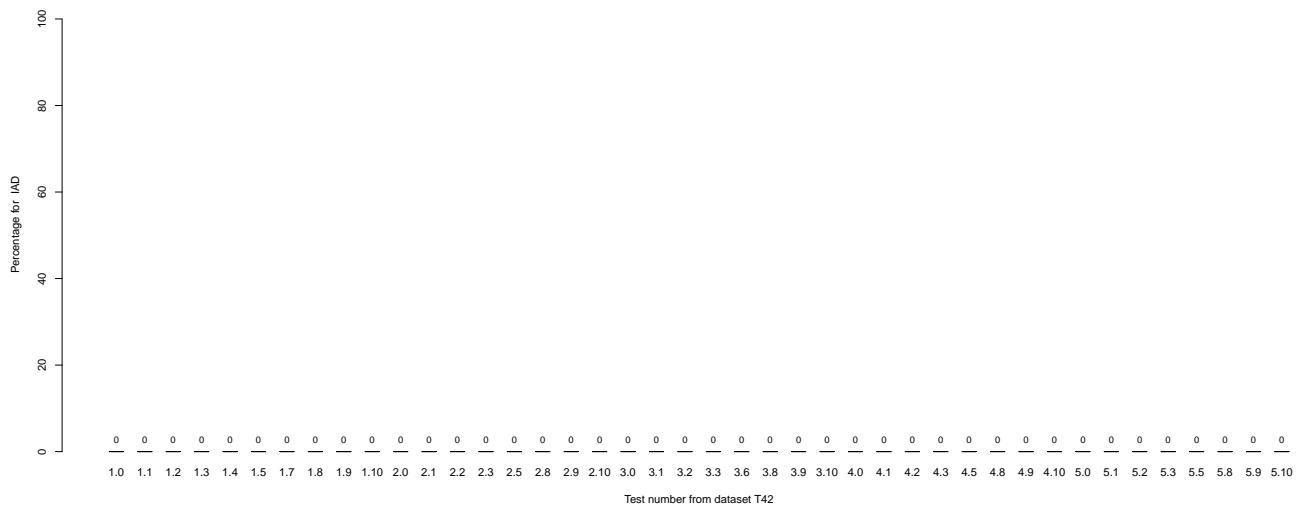


Fig. 23. Quality results for FRBR-ML with dataset T42 and metric IAD

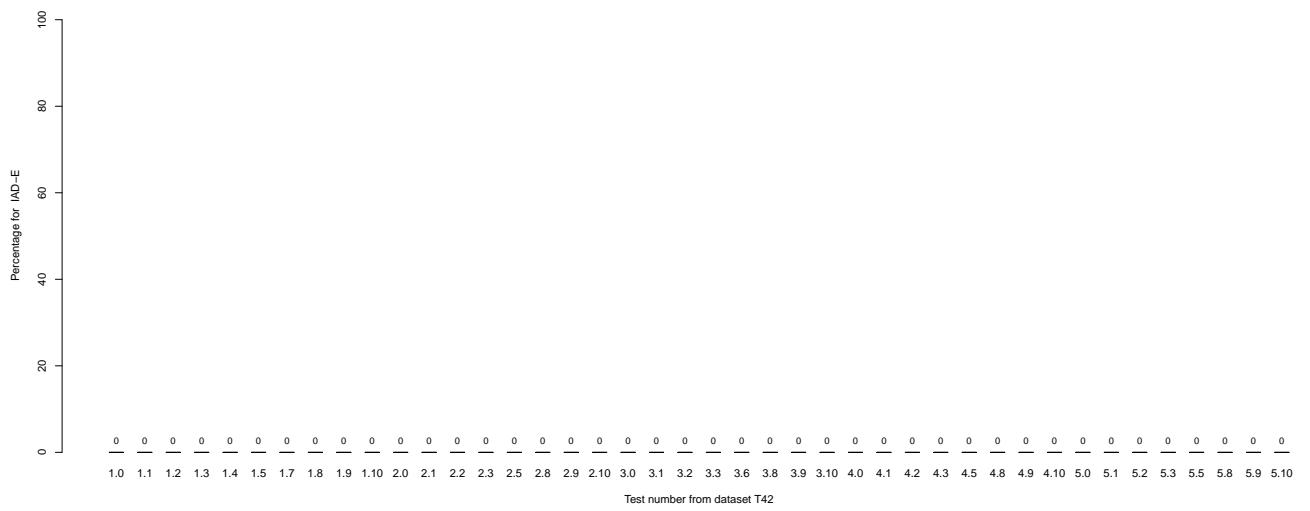


Fig. 24. Quality results for FRBR-ML with dataset T42 and metric IAD-E

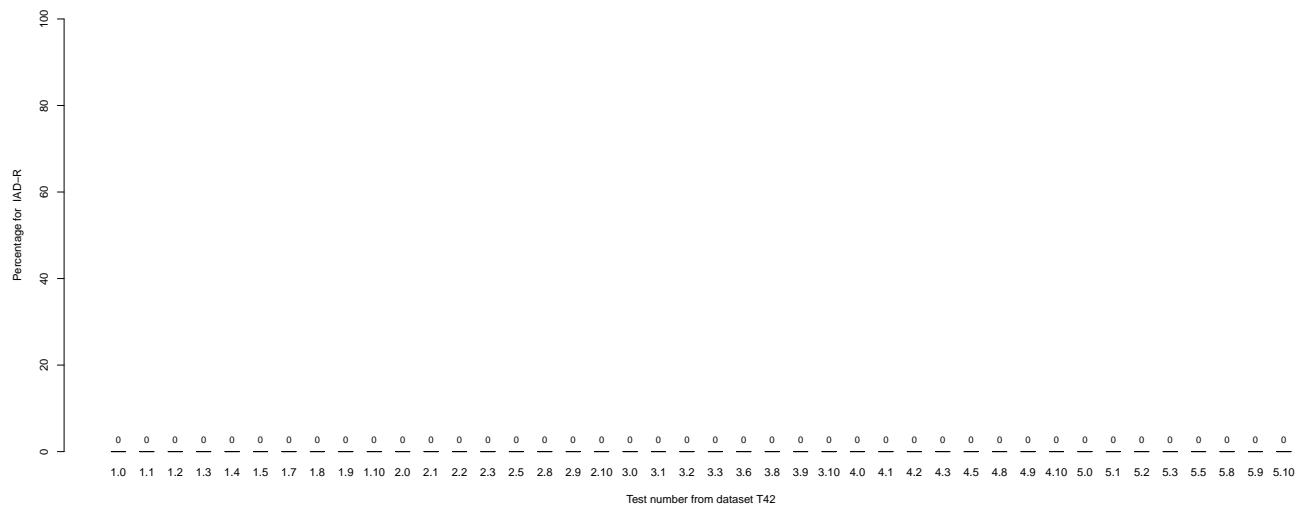


Fig. 25. Quality results for FRBR-ML with dataset T42 and metric IAD-R

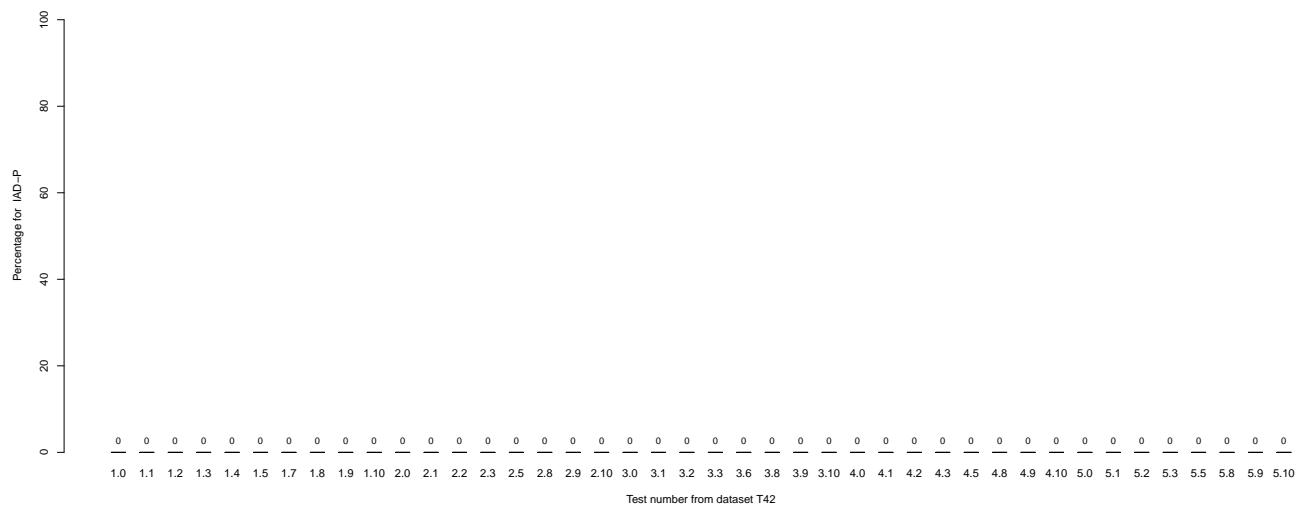


Fig. 26. Quality results for FRBR-ML with dataset T42 and metric IAD-P

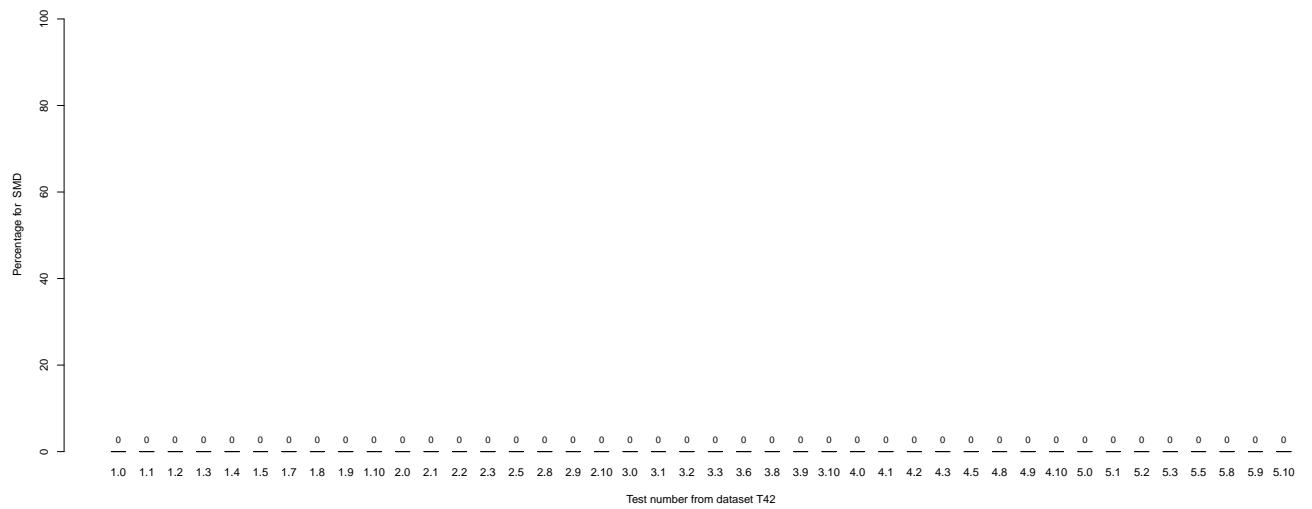


Fig. 27. Quality results for FRBR-ML with dataset T42 and metric SMD

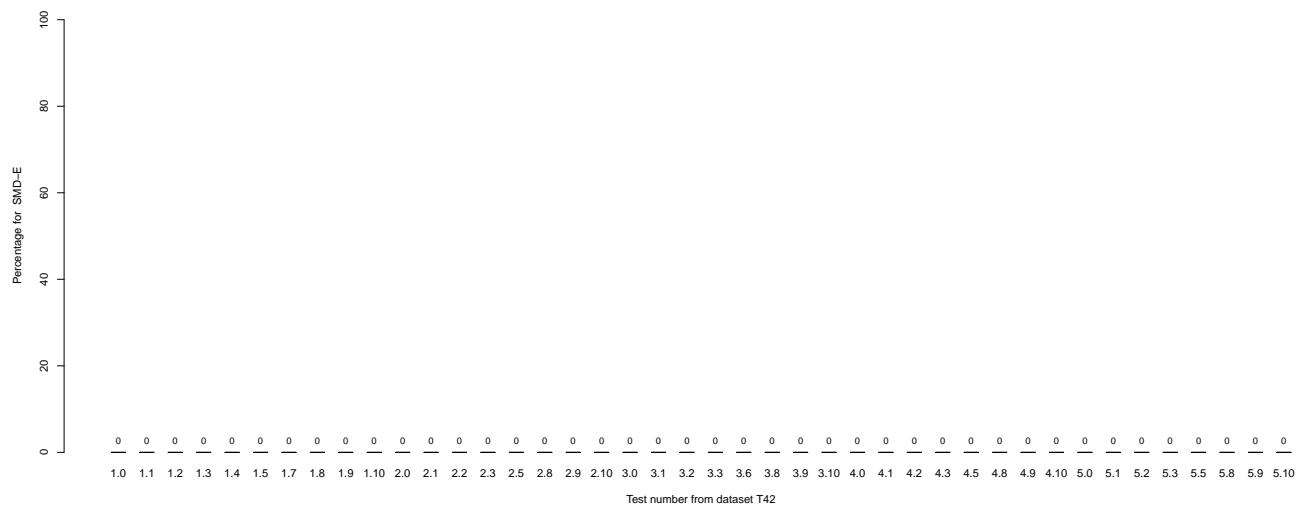


Fig. 28. Quality results for FRBR-ML with dataset T42 and metric SMD-E

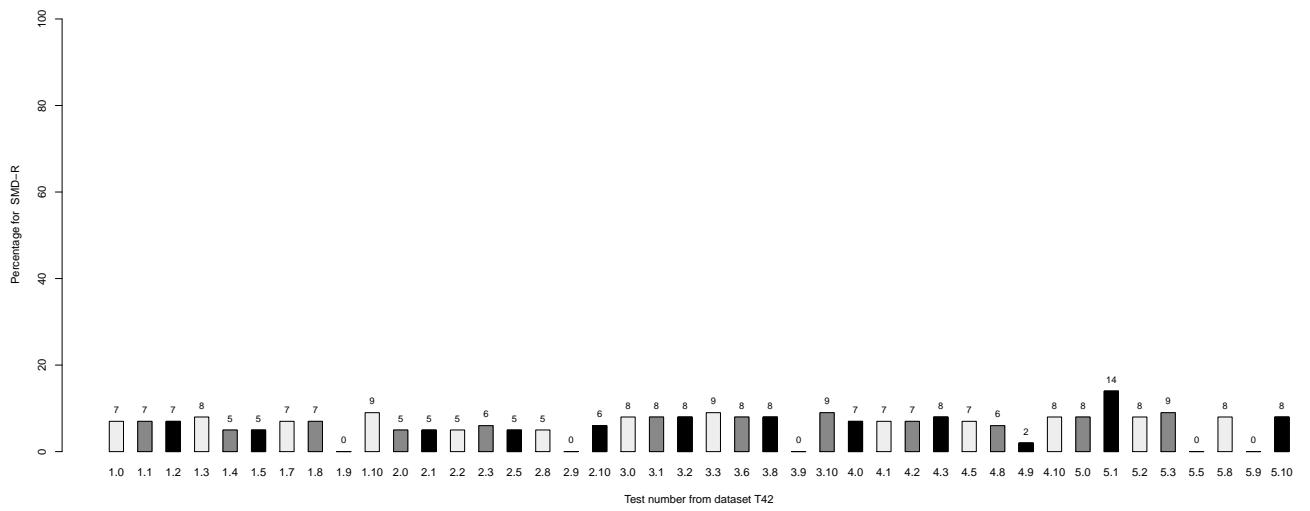


Fig. 29. Quality results for FRBR-ML with dataset T42 and metric SMD-R

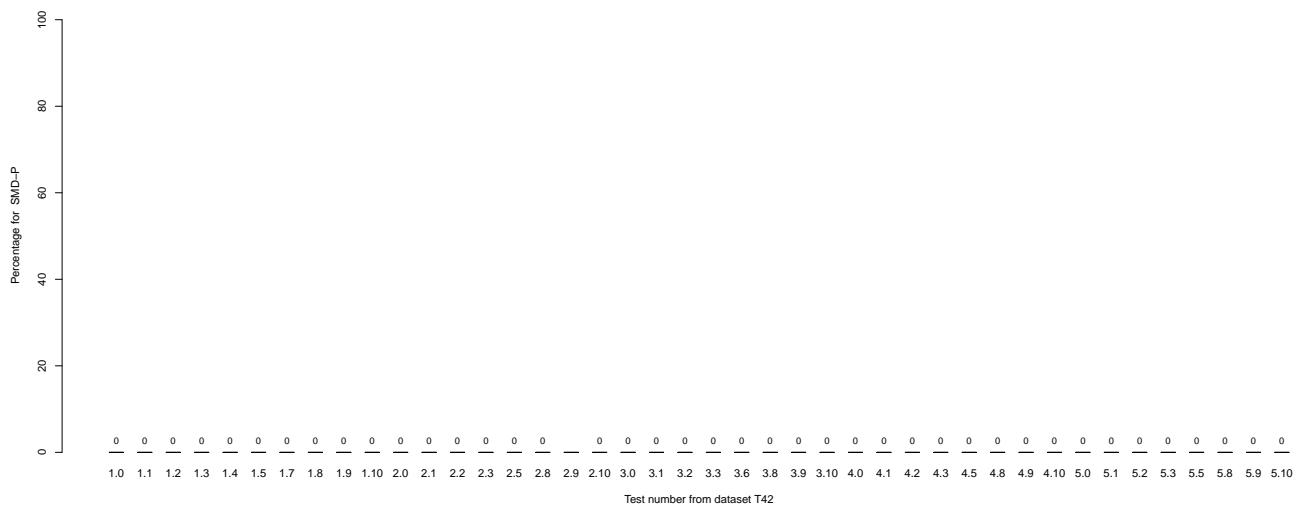


Fig. 30. Quality results for FRBR-ML with dataset T42 and metric SMD-P

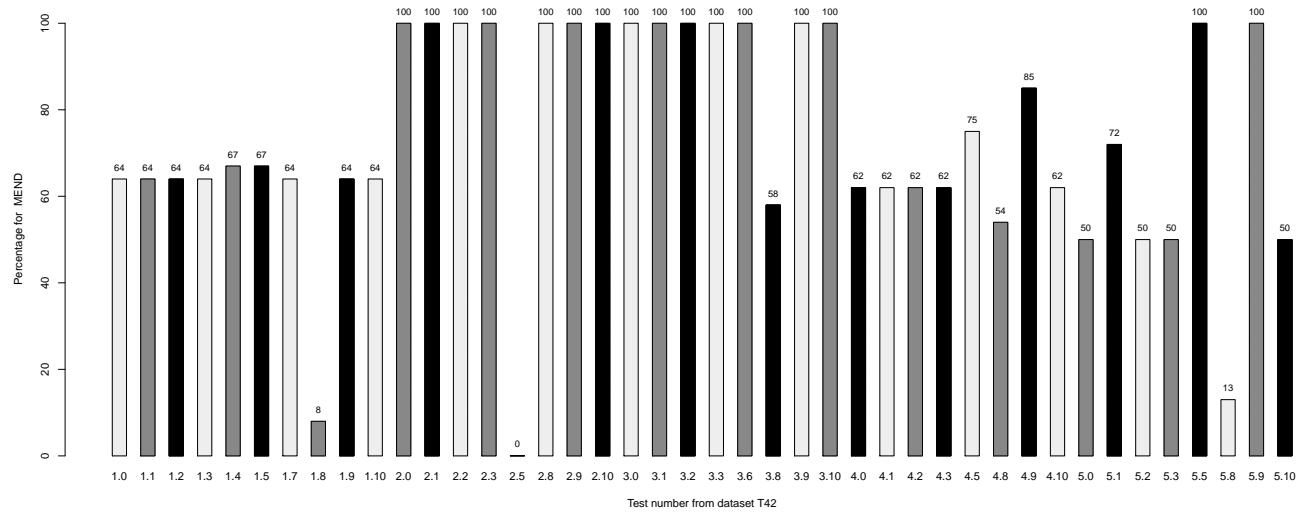


Fig. 31. Quality results for Variation/VFRBR with dataset T42 and metric MEND

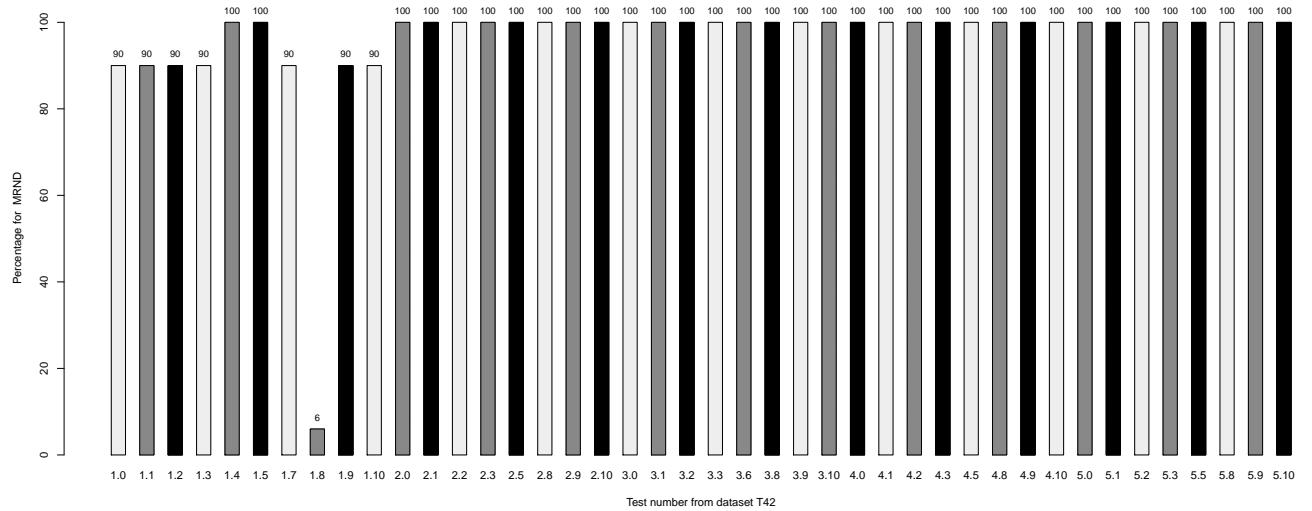


Fig. 32. Quality results for Variation/VFRBR with dataset T42 and metric MRND

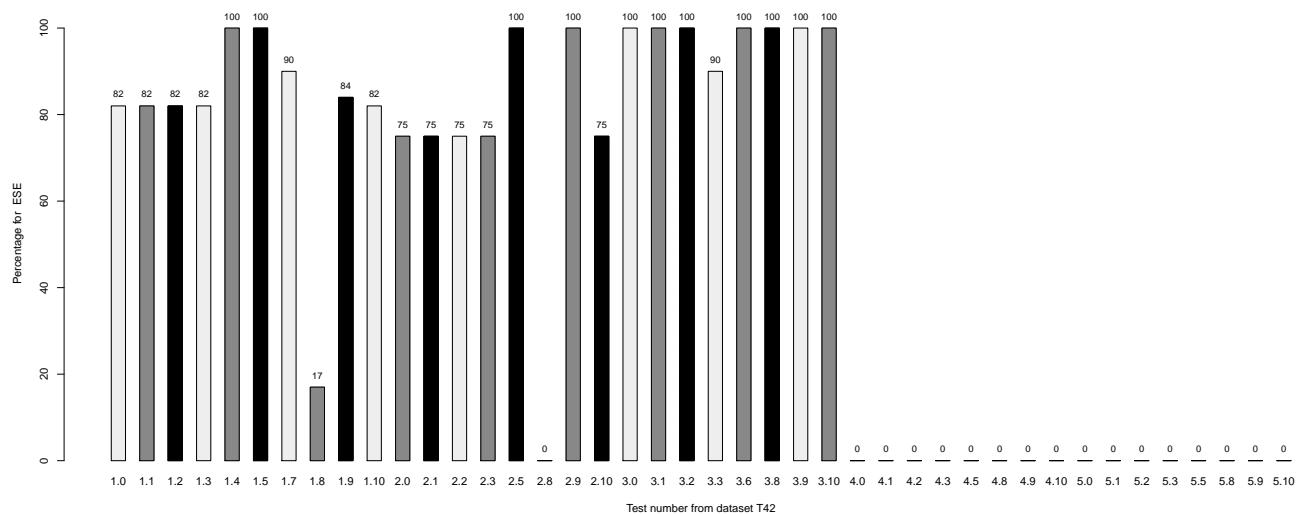


Fig. 33. Quality results for Variation/VFRBR with dataset T42 and metric ESE

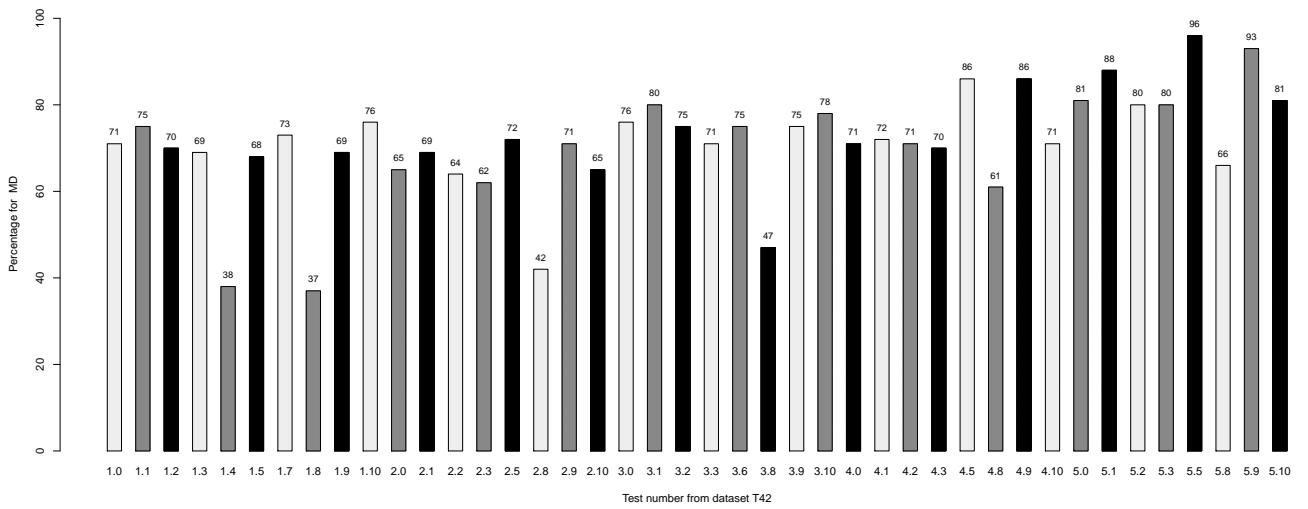


Fig. 34. Quality results for Variation/VFRBR with dataset T42 and metric MD

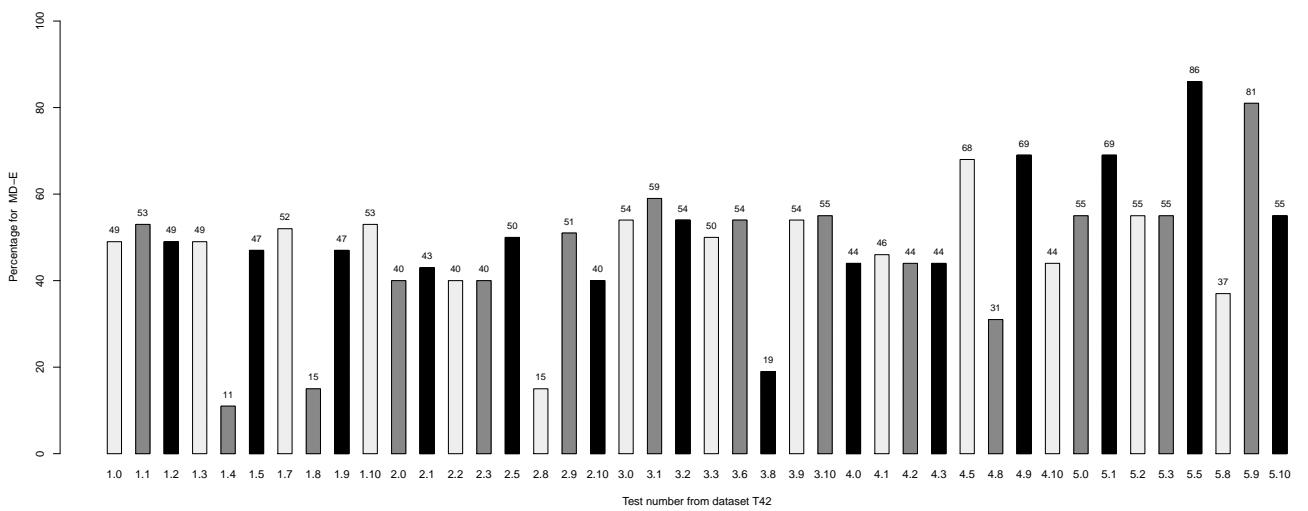


Fig. 35. Quality results for Variation/VFRBR with dataset T42 and metric MD-E

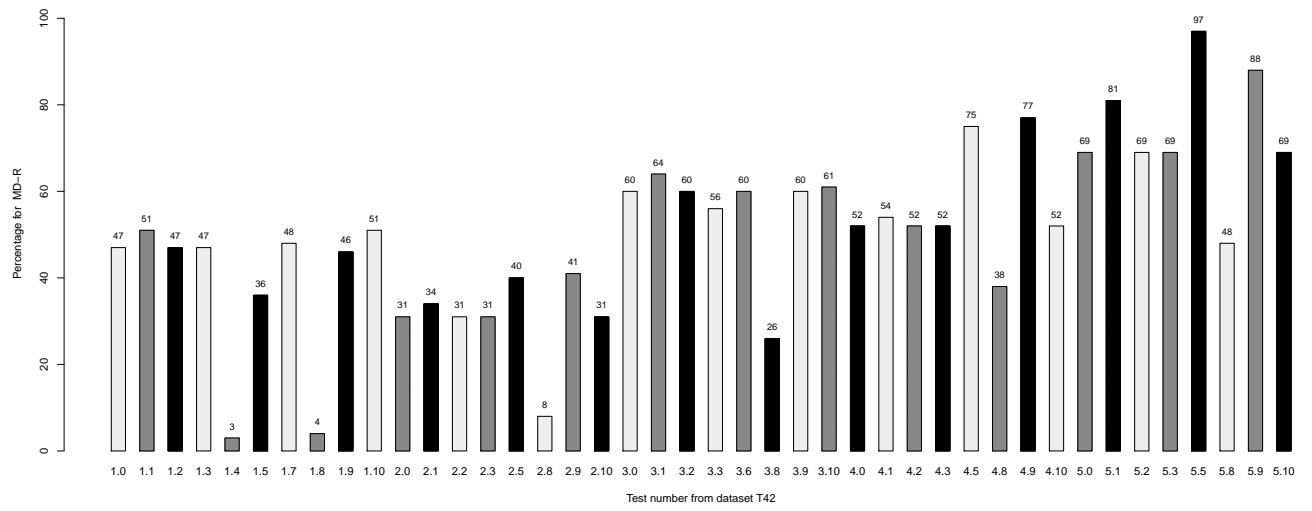


Fig. 36. Quality results for Variation/VFRBR with dataset T42 and metric MD-R

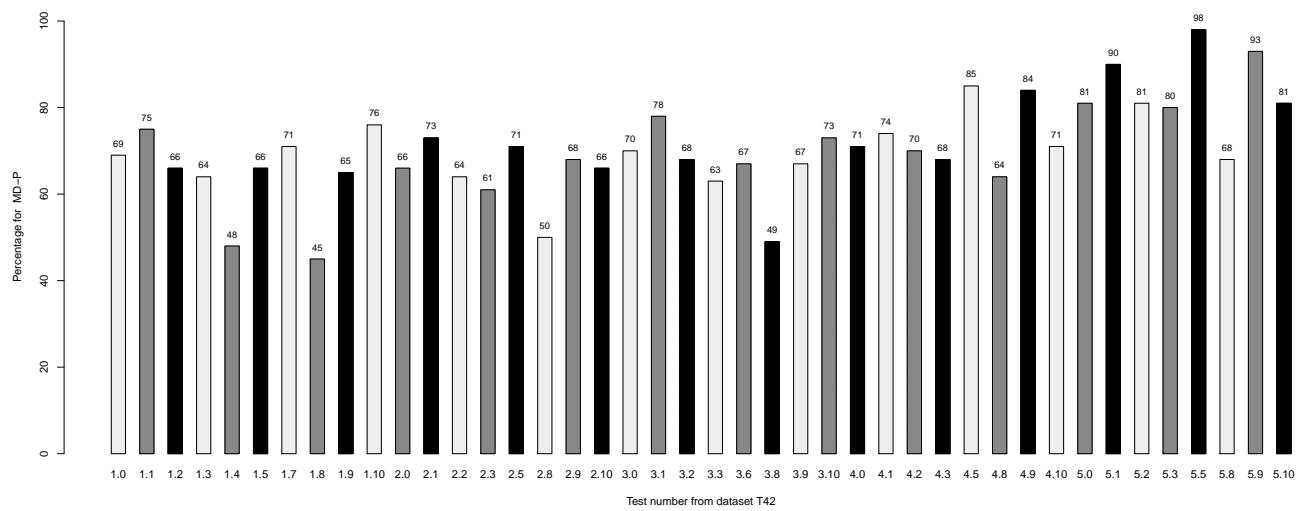


Fig. 37. Quality results for Variation/VFRBR with dataset T42 and metric MD-P

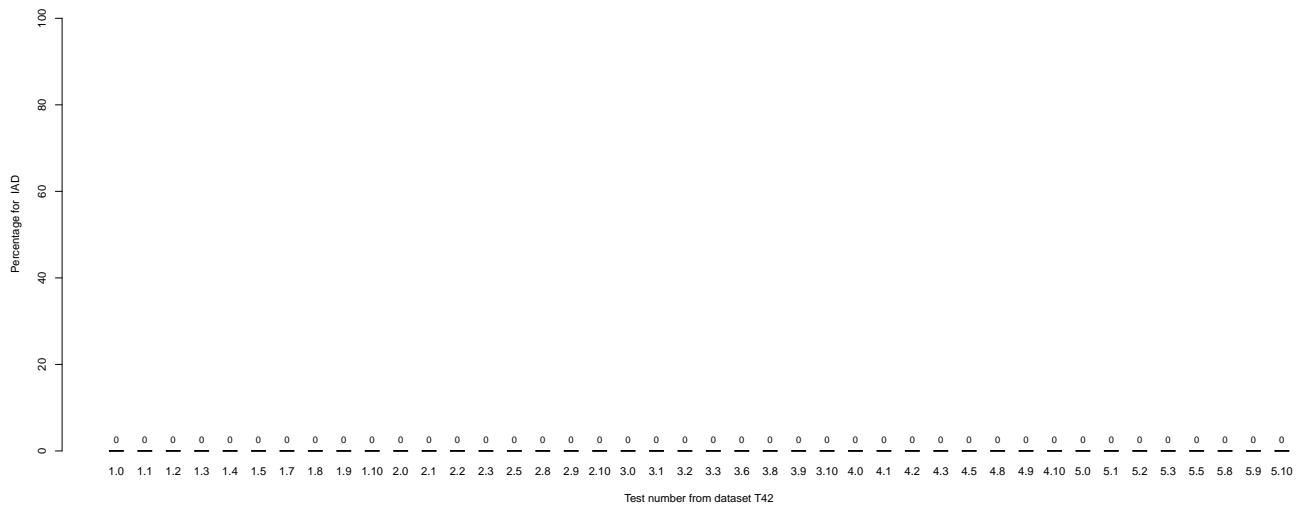


Fig. 38. Quality results for Variation/VFRBR with dataset T42 and metric IAD

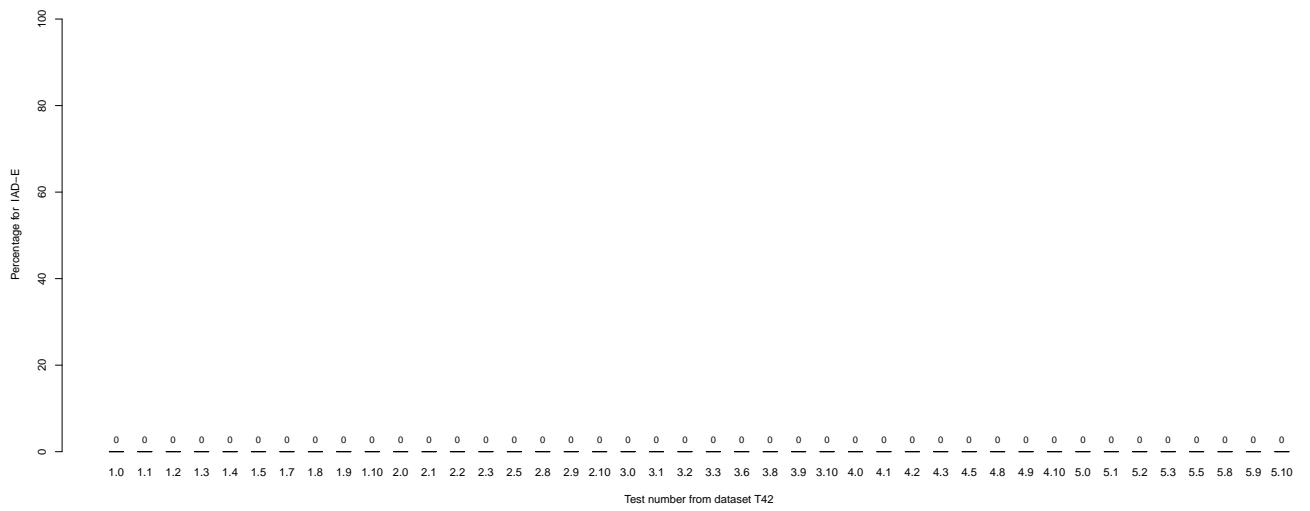


Fig. 39. Quality results for Variation/VFRBR with dataset T42 and metric IAD-E

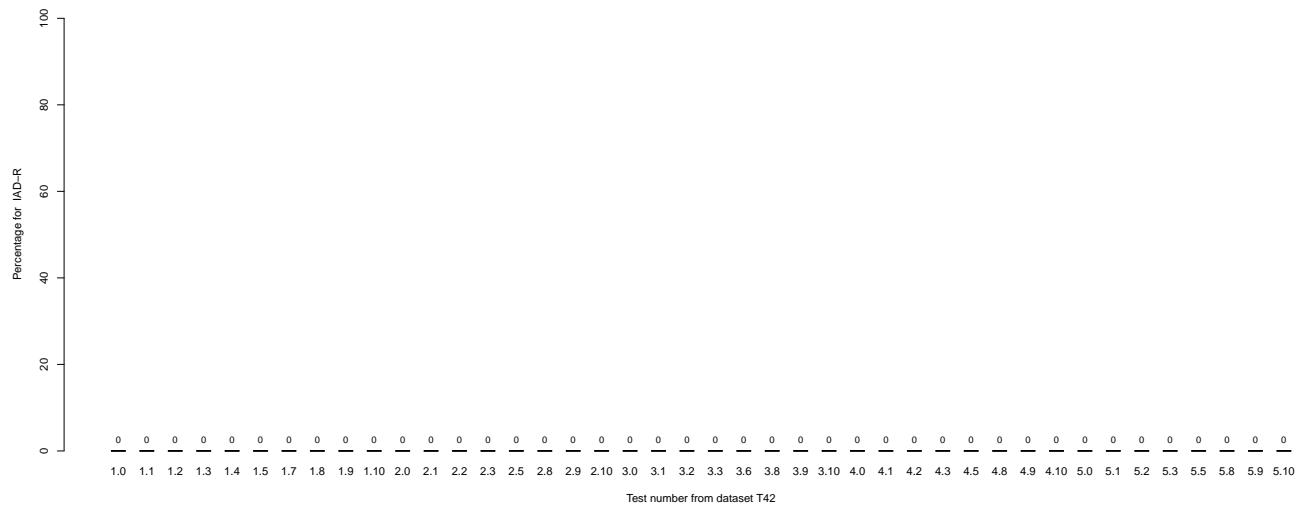


Fig. 40. Quality results for Variation/VFRBR with dataset T42 and metric IAD-R

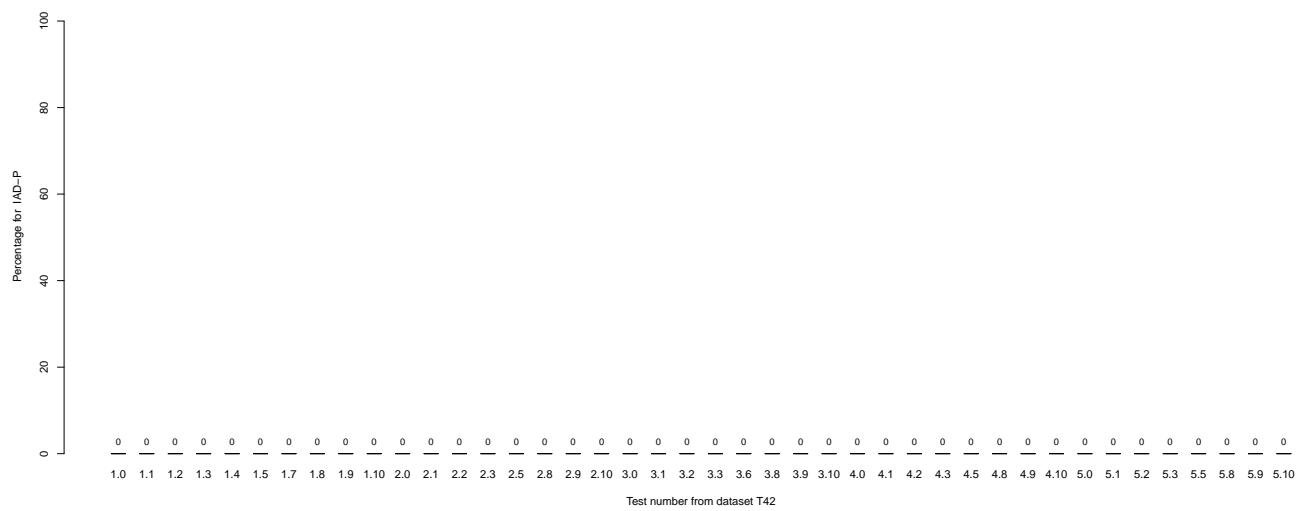


Fig. 41. Quality results for Variation/VFRBR with dataset T42 and metric IAD-P

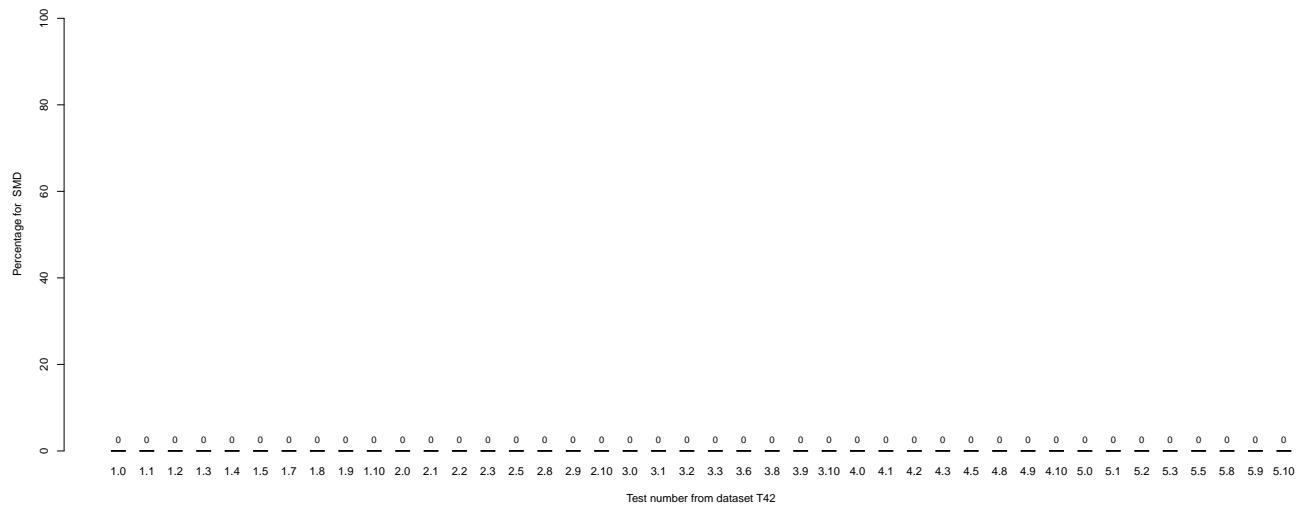


Fig. 42. Quality results for Variation/VFRBR with dataset T42 and metric SMD

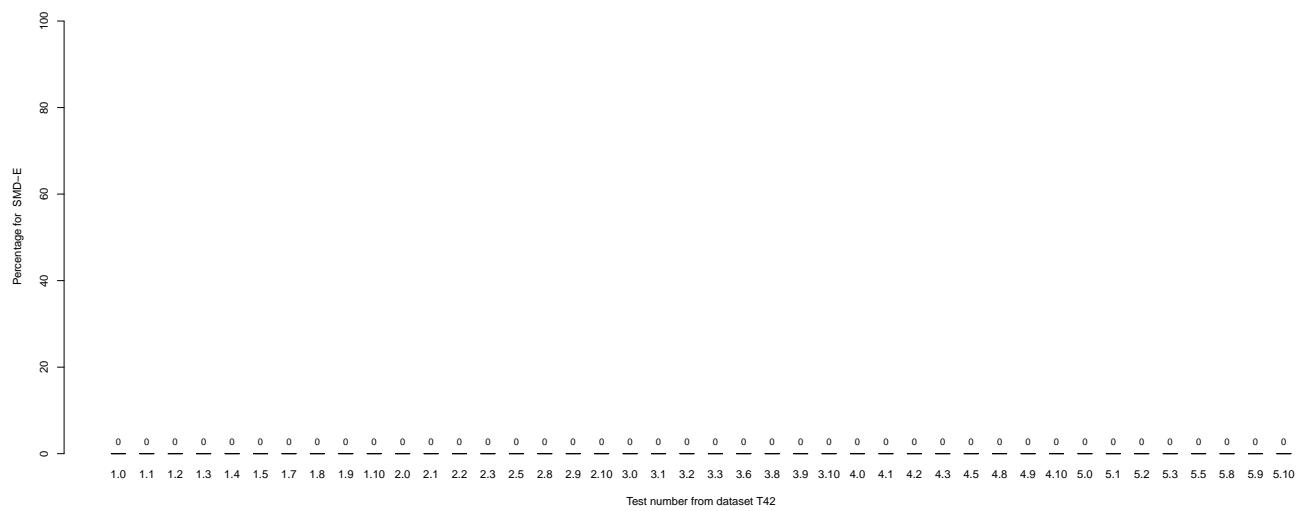


Fig. 43. Quality results for Variation/VFRBR with dataset T42 and metric SMD-E

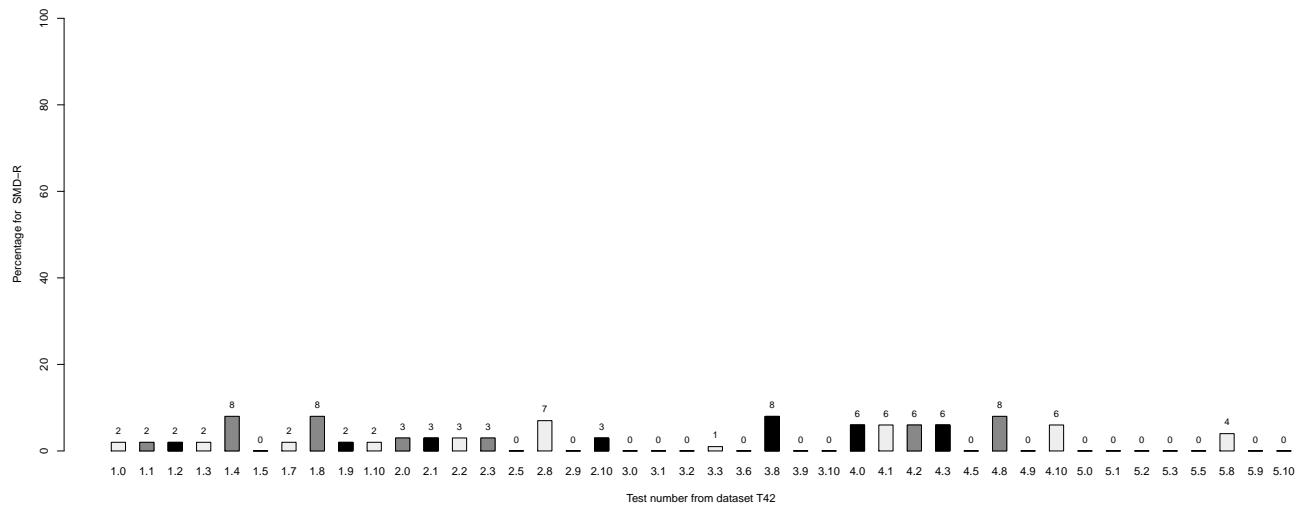


Fig. 44. Quality results for Variation/VFRBR with dataset T42 and metric SMD-R

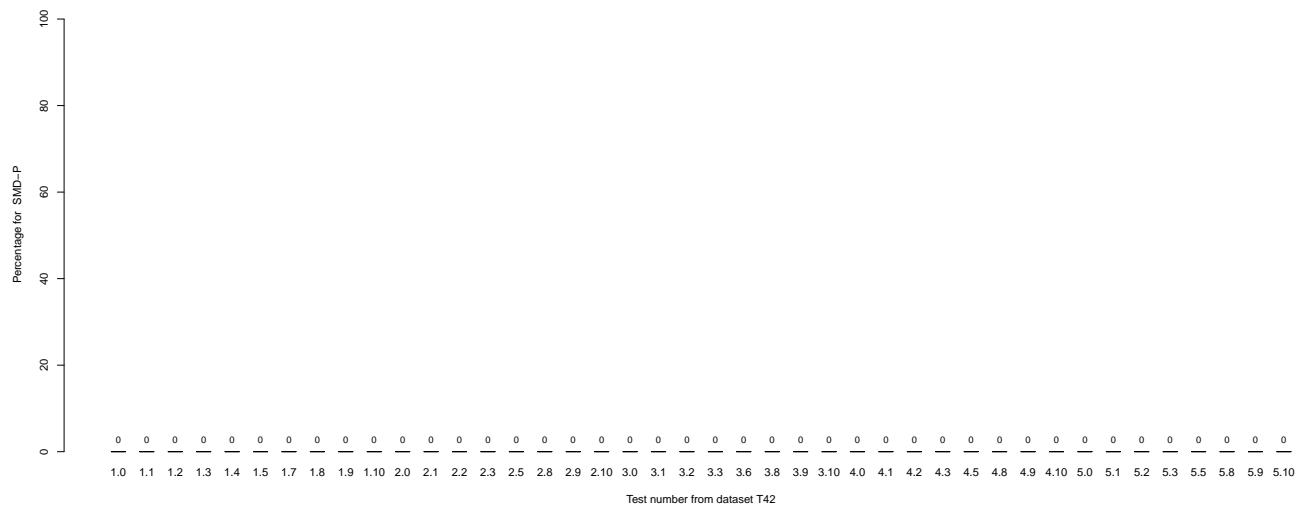


Fig. 45. Quality results for Variation/VFRBR with dataset T42 and metric SMD-P

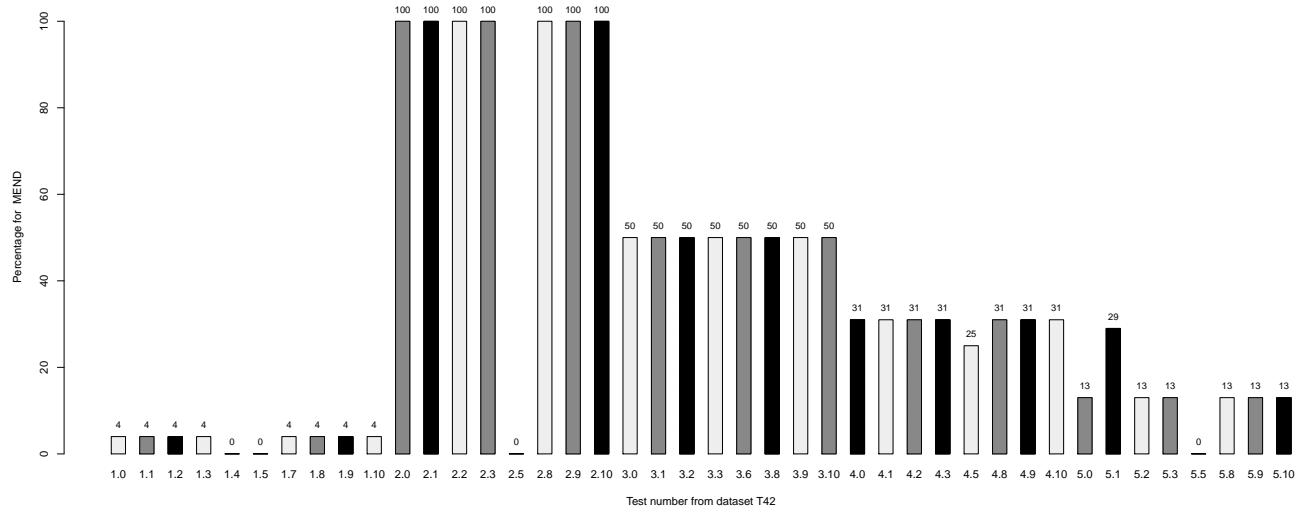


Fig. 46. Quality results for XC with dataset T42 and metric MEND

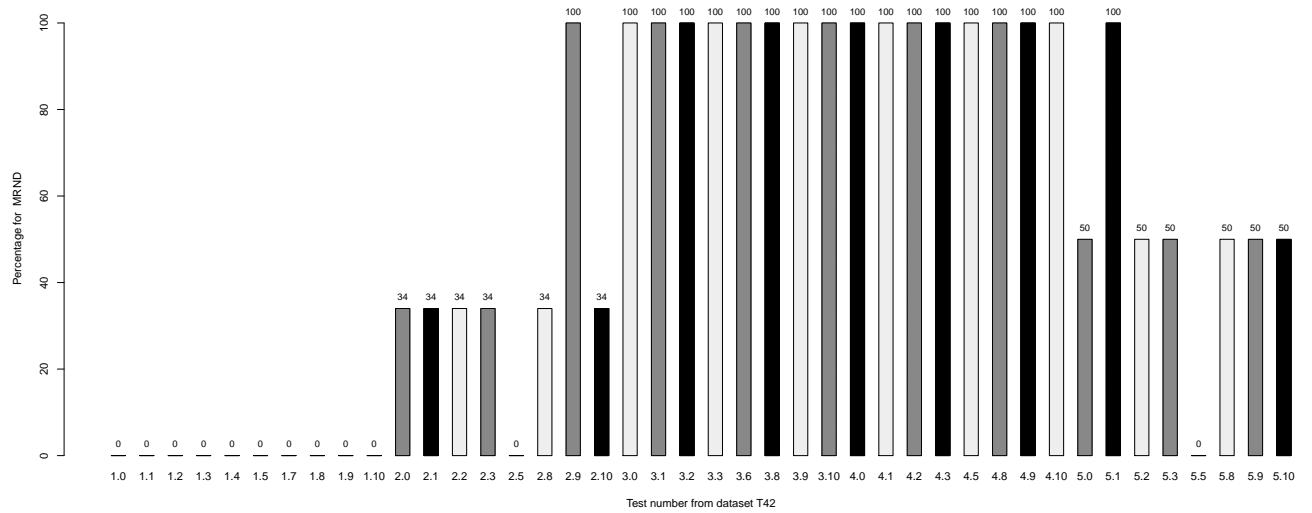


Fig. 47. Quality results for XC with dataset T42 and metric MRND

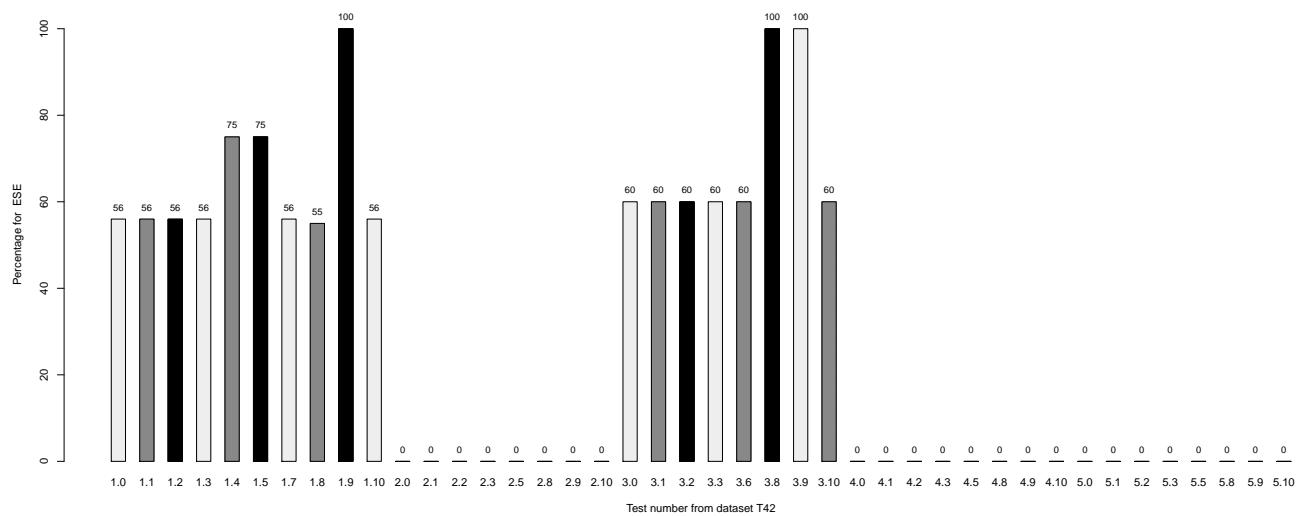


Fig. 48. Quality results for XC with dataset T42 and metric ESE

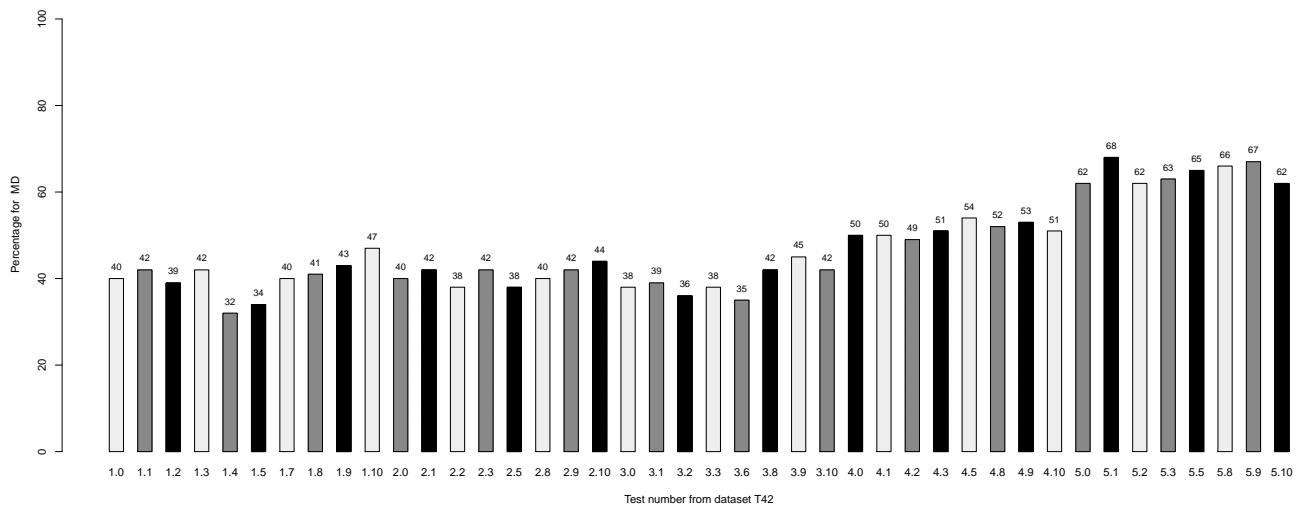


Fig. 49. Quality results for XC with dataset T42 and metric MD

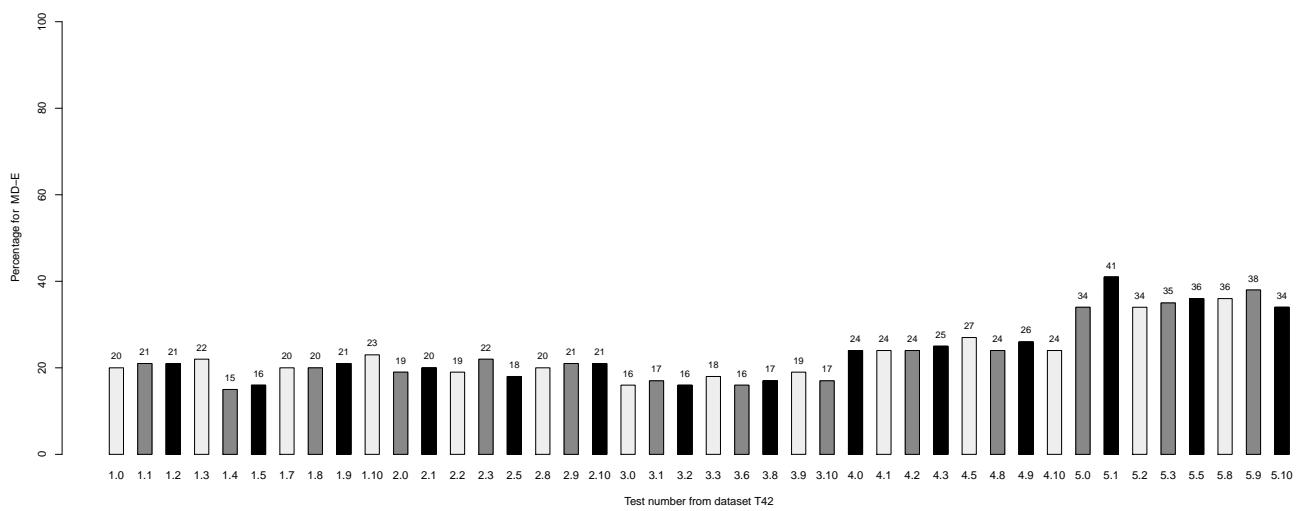


Fig. 50. Quality results for XC with dataset T42 and metric MD-E

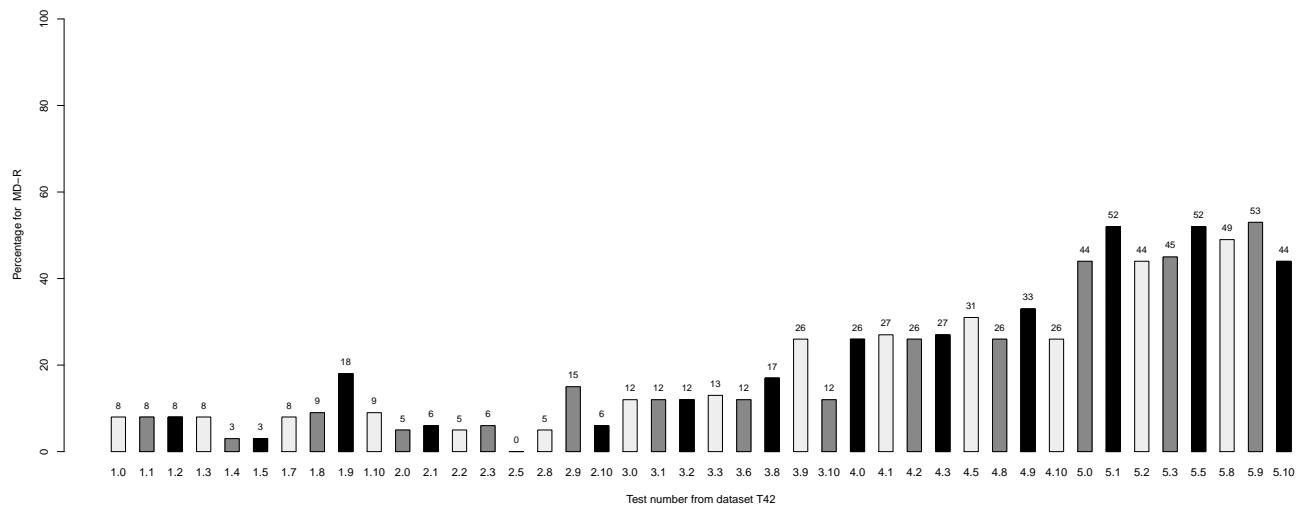


Fig. 51. Quality results for XC with dataset T42 and metric MD-R

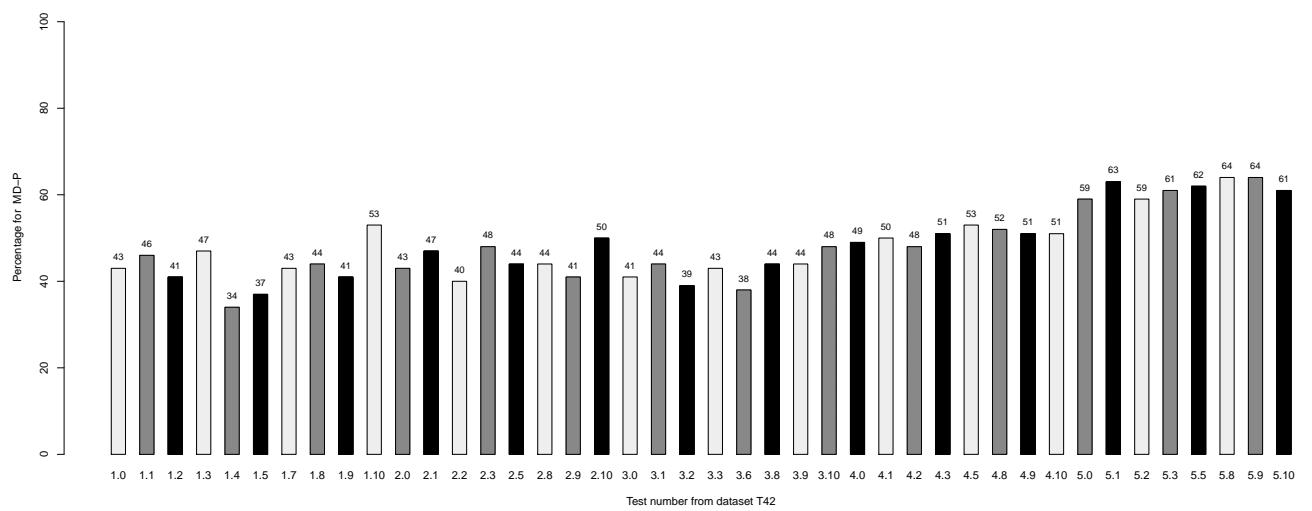


Fig. 52. Quality results for XC with dataset T42 and metric MD-P

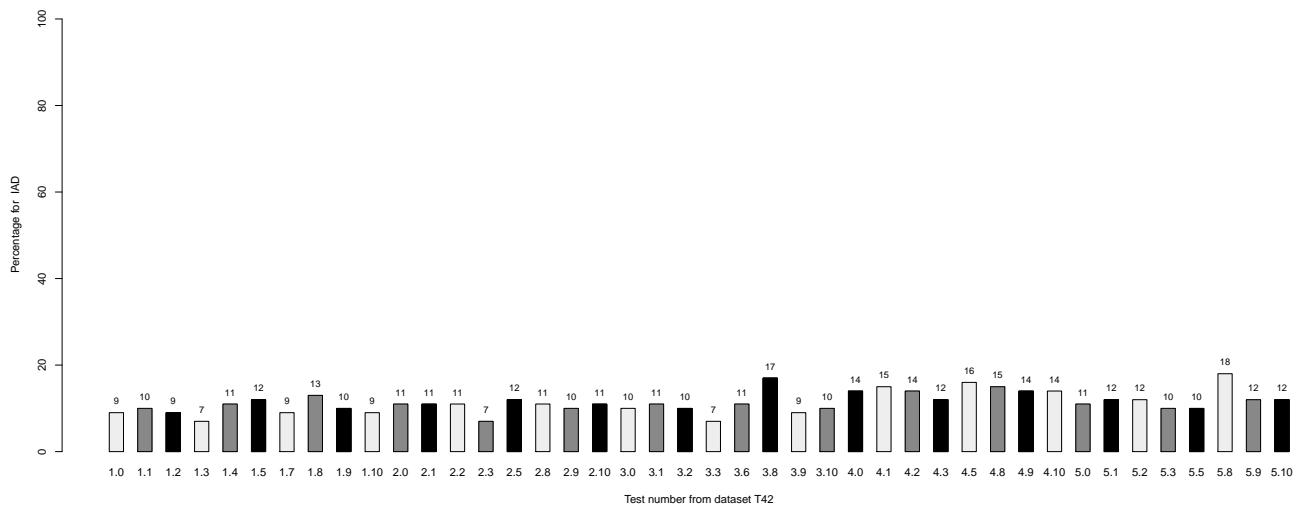


Fig. 53. Quality results for XC with dataset T42 and metric IAD

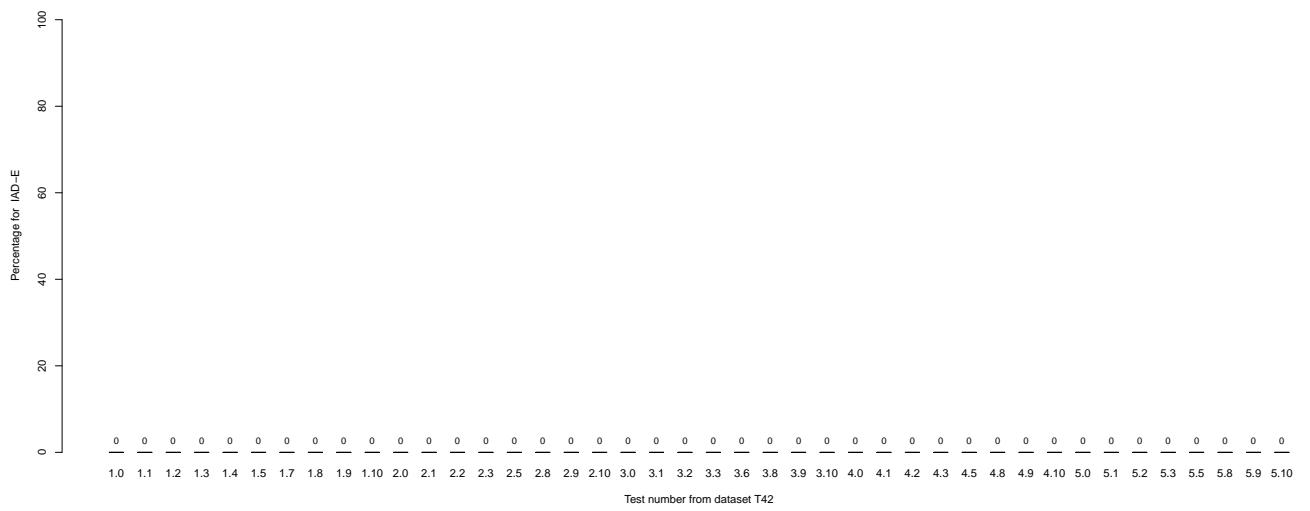


Fig. 54. Quality results for XC with dataset T42 and metric IAD-E

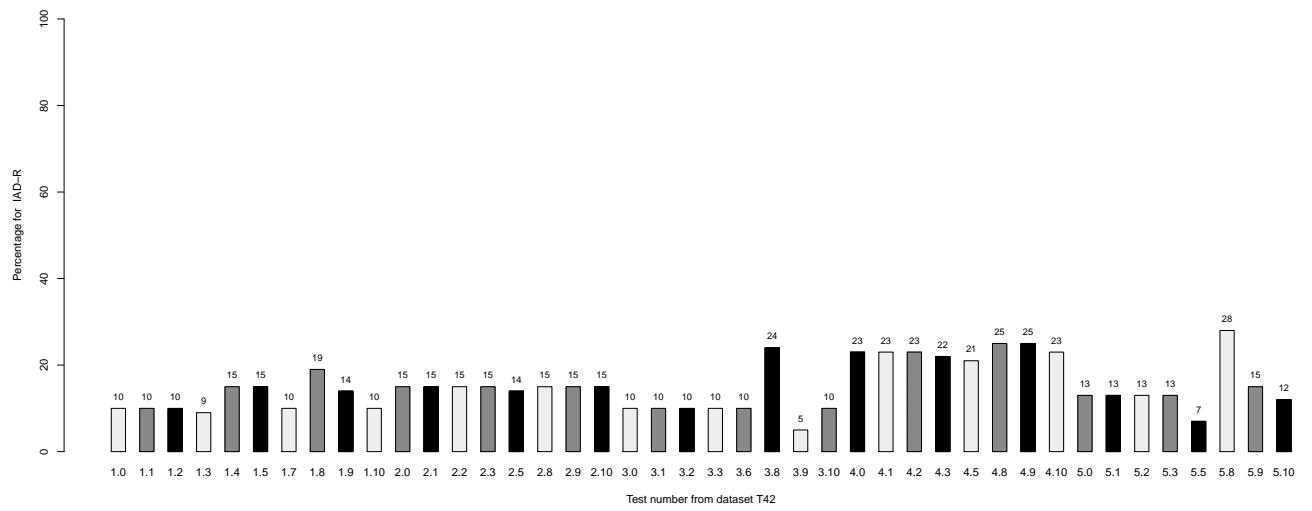


Fig. 55. Quality results for XC with dataset T42 and metric IAD-R

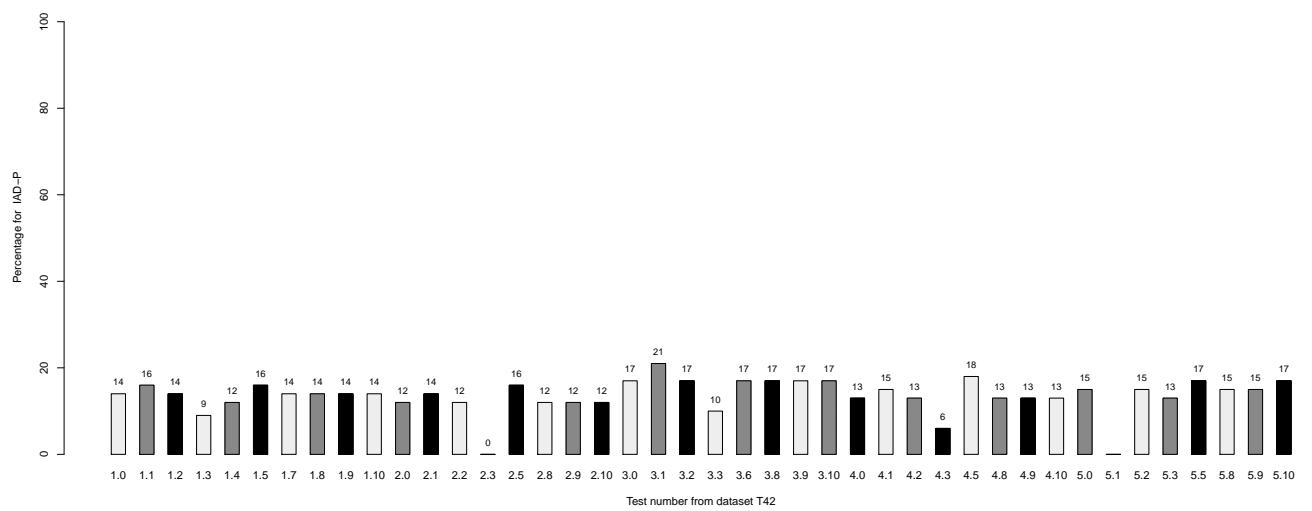


Fig. 56. Quality results for XC with dataset T42 and metric IAD-P

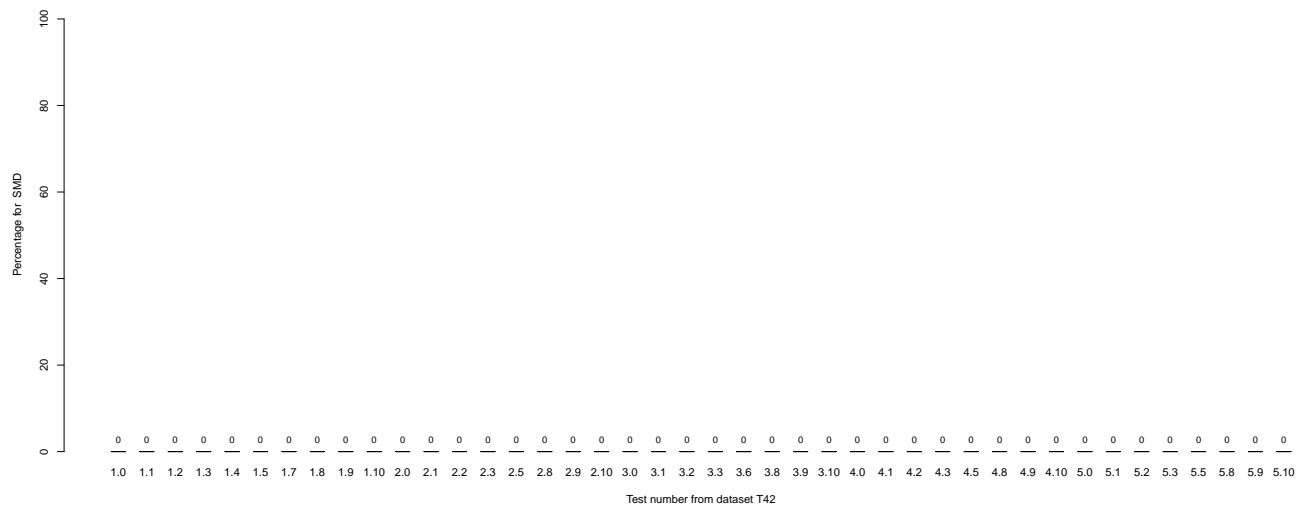


Fig. 57. Quality results for XC with dataset T42 and metric SMD

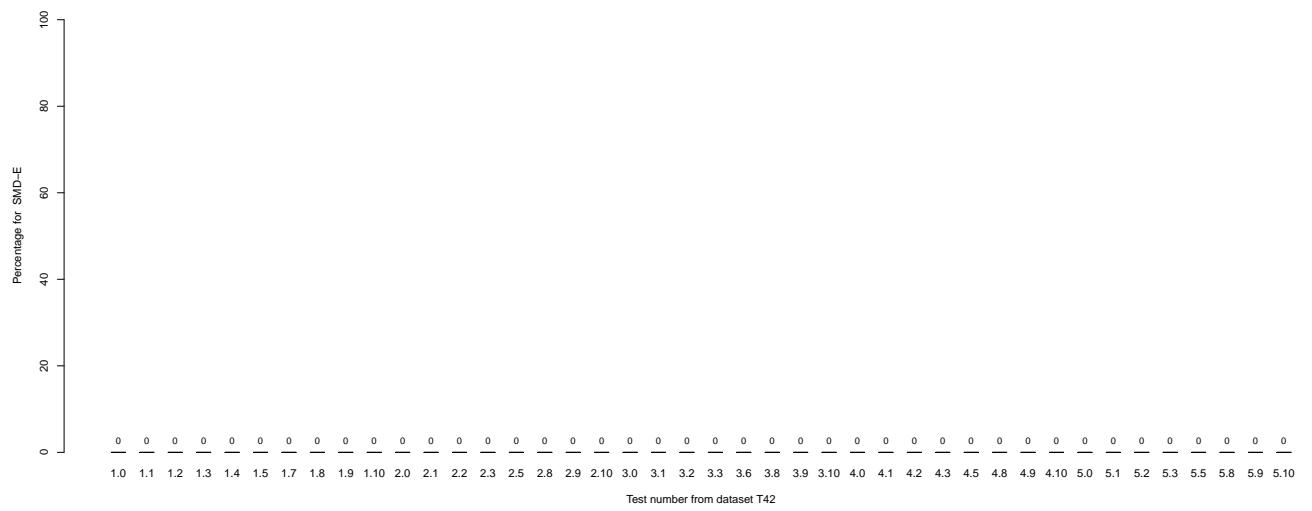


Fig. 58. Quality results for XC with dataset T42 and metric SMD-E

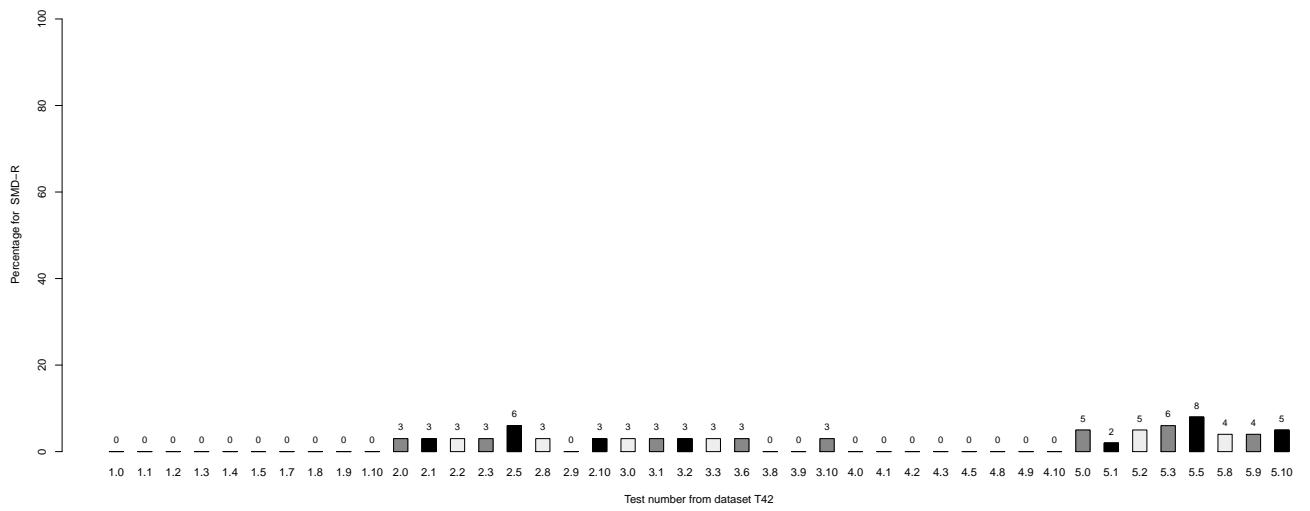


Fig. 59. Quality results for XC with dataset T42 and metric SMD-R

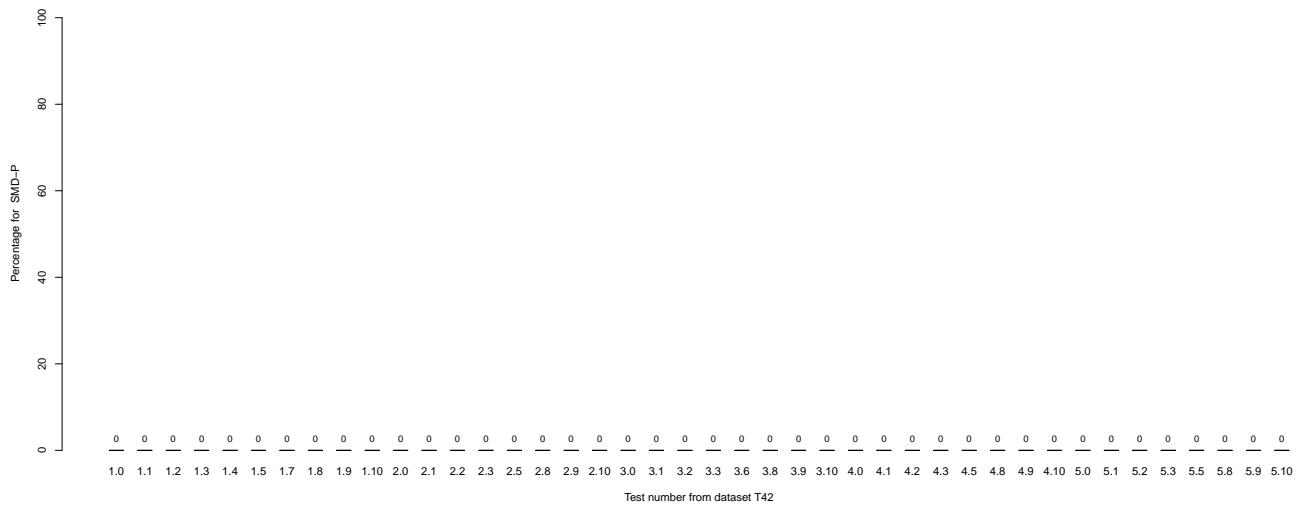


Fig. 60. Quality results for XC with dataset and metric SMD-P

5 Comparing tools in real-world context

The objective of this second experiment is to compare FRBRization tools in a real-world context using both FRBRization and post-FRBRization metrics. All tools rely on their basic set of rules (no tuning). Contrary to the dataset T42, all records from the dataset BIB-RCAT come from various institutions. This means that these real-world records can be rather simple or may contain several bibliographic patterns and issues. Note that the FRBRization metrics ETC, ETD and NRT could not be presented, because the tools did not implement these features. Instead, we provide the overall execution time for FRBRizing the dataset BIB-RCAT. The post-FRBRization metric DLE is also not given, since there exists many authority files or knowledge bases (e.g., Linked Open Data) and the expert FRBRized collection cannot include a link for each of these sources.

In the following plots, the results are organized by tool.

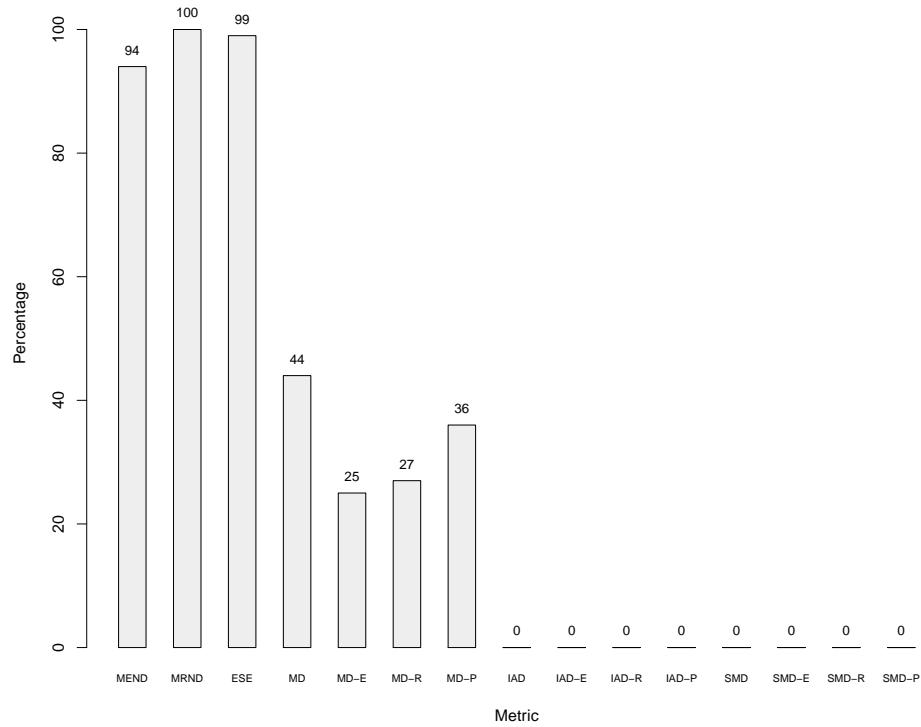


Fig. 61. Quality results for FRBR-ML basic (no tuning) on dataset BIB-RCAT

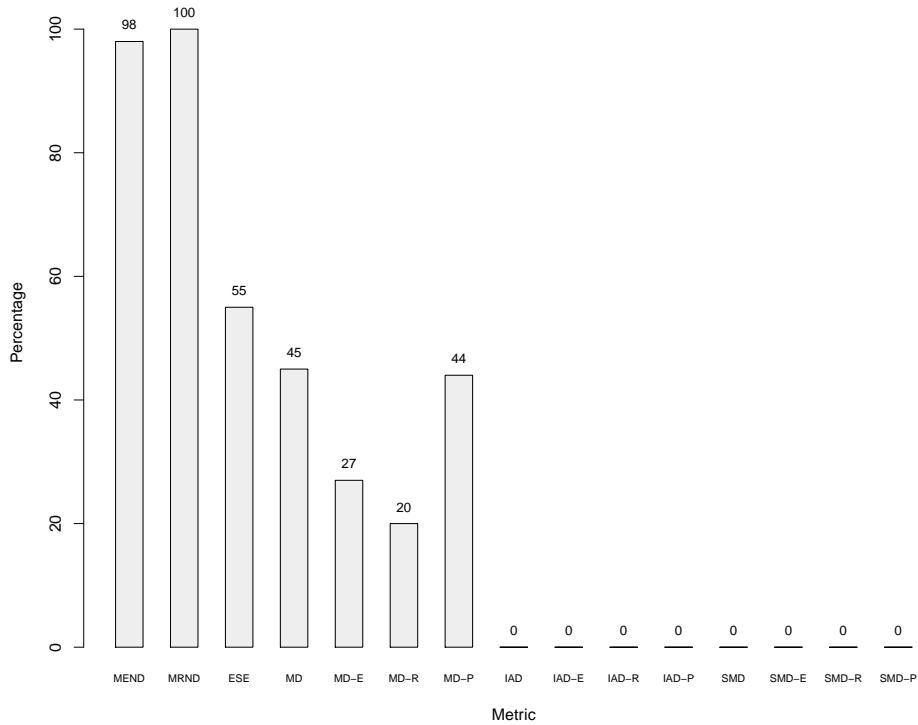


Fig. 62. Quality results for VFRBR on dataset BIB-RCAT

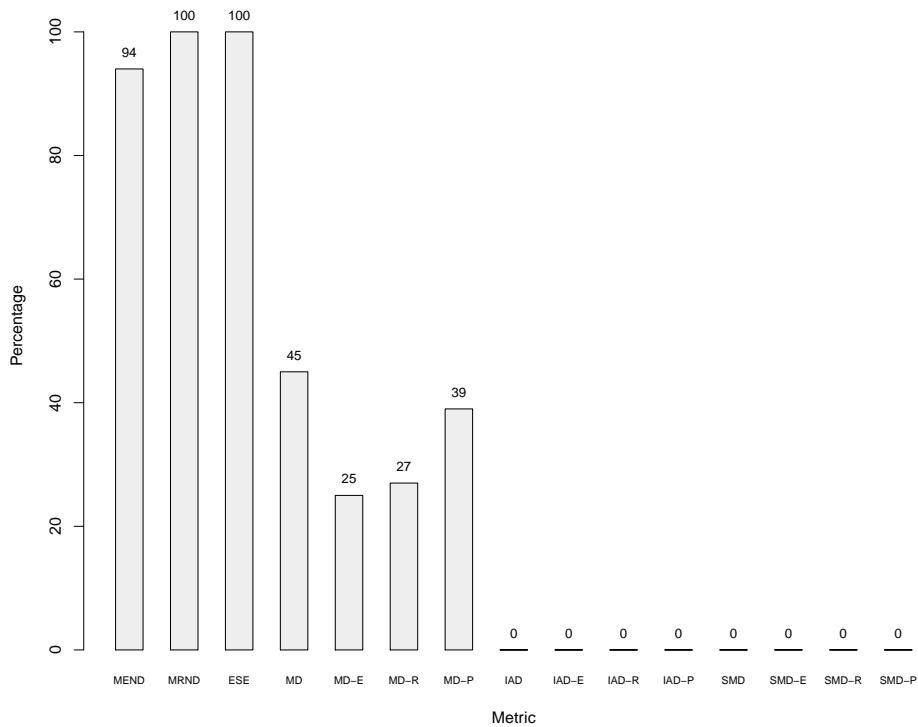


Fig. 63. Quality results for XC on dataset BIB-RCAT

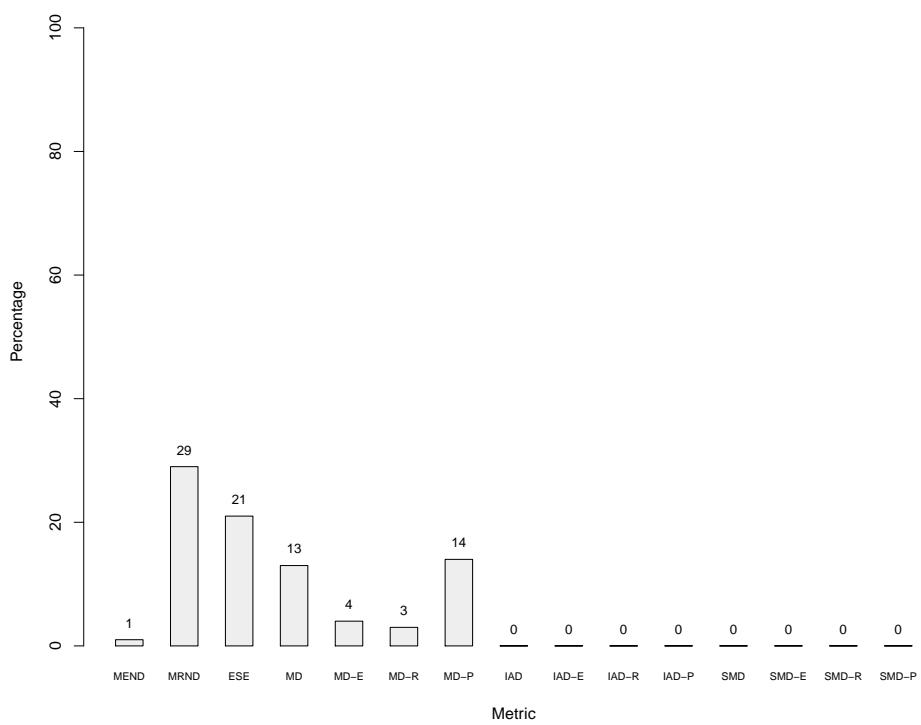


Fig. 64. Quality results for FRBR-ML tuned (enhanced set of rules) on dataset BIB-RCAT

6 Additional information about experiments

In this section, we provide screenshots of the tools that have been used during experiments.

Figure 65 depicts (an extract of) the FRBRized collection of the BIB-RCAT dataset using FRBR-ML.

FRBR-ML
Open the home page file:///Users/takhirov/Dropbox/projects/marc2frbr/examples/marc21/output/frbr.example.html Naimdjou

FRBR-ML v0.1

This is a user-friendly display of XML file: /Users/takhirov/Dropbox/projects/FRBR-ML/examples/marc21/output/frbr.example.records.xml

Mireille l'abeille (Krings, Antoon)
Mireille l'abeille
Antoon Krings / Mireille l'abeille

Un sac de billes (Joffo, Joseph)
Un sac de billes
Joseph Joffo; illustrations Claude Lapointe / Un sac de billes
Joseph Joffo; illustrations Claude Lapointe / Un sac de billes

La Vénus d'Ille (Mérimée, Prosper)
La Vénus d'Ille ; La chambre bleue La Vénus d'Ille
Prosper Mérimée; Prosper Mérimée / La Vénus d'Ille ; La chambre bleue
Mérimée; édition Dominique Fleur-Schluthess, Claudine Zenou-Grinstein / La Vénus d'Ille

Les allumettes suédoises (Sabatier, Robert)
Les allumettes suédoises
Robert Sabatier; lu par Jean Barrier / Les allumettes suédoises
Robert Sabatier / Les allumettes suédoises
Robert Sabatier; illustrations Louis Constantin / Les allumettes suédoises
Robert Sabatier; illustrations Louis Constantin / Les allumettes suédoises
Robert Sabatier; illustrations Louis Constantin / Les allumettes suédoises

Tous à poil ! (Franek, Claire)
Tous à poil !
Claire Franek, Marc Daniau / Tous à poil !

Les contes bleus du chat perché (Aymé, Marcel)
Les contes bleus du chat perché
Marcel Aymé; illustrations Claudine et Roland Sabatier / Les contes bleus du chat perché

Un sac de billes (Bouton, Alain)
Un sac de billes
d'après Joseph Joffo; scénario Alain Bouton; dessins Marc Malès / Un sac de billes

365 contes pour tous les âges (Bloch, Muriel)
365 contes pour tous les âges
Muriel Bloch; illustrations Mireille Vautier / 365 contes pour tous les âges
Muriel Bloch; illustrations Grégoire Solotareff / 365 contes pour tous les âges
Muriel Bloch / 365 contes pour tous les âges

La belle lisse poire du prince de Motordu ()
La belle lisse poire du prince de Motordu : et 4 autres histoires La belle lisse poire du

Fig. 65. Extract of the FRBRized BIB-RCAT dataset in FRBR-ML

Figure 66 depicts (an extract of) the relationships between FRBR entities in the FRBRized collection of the BIB-RCAT dataset using FRBR-ML.

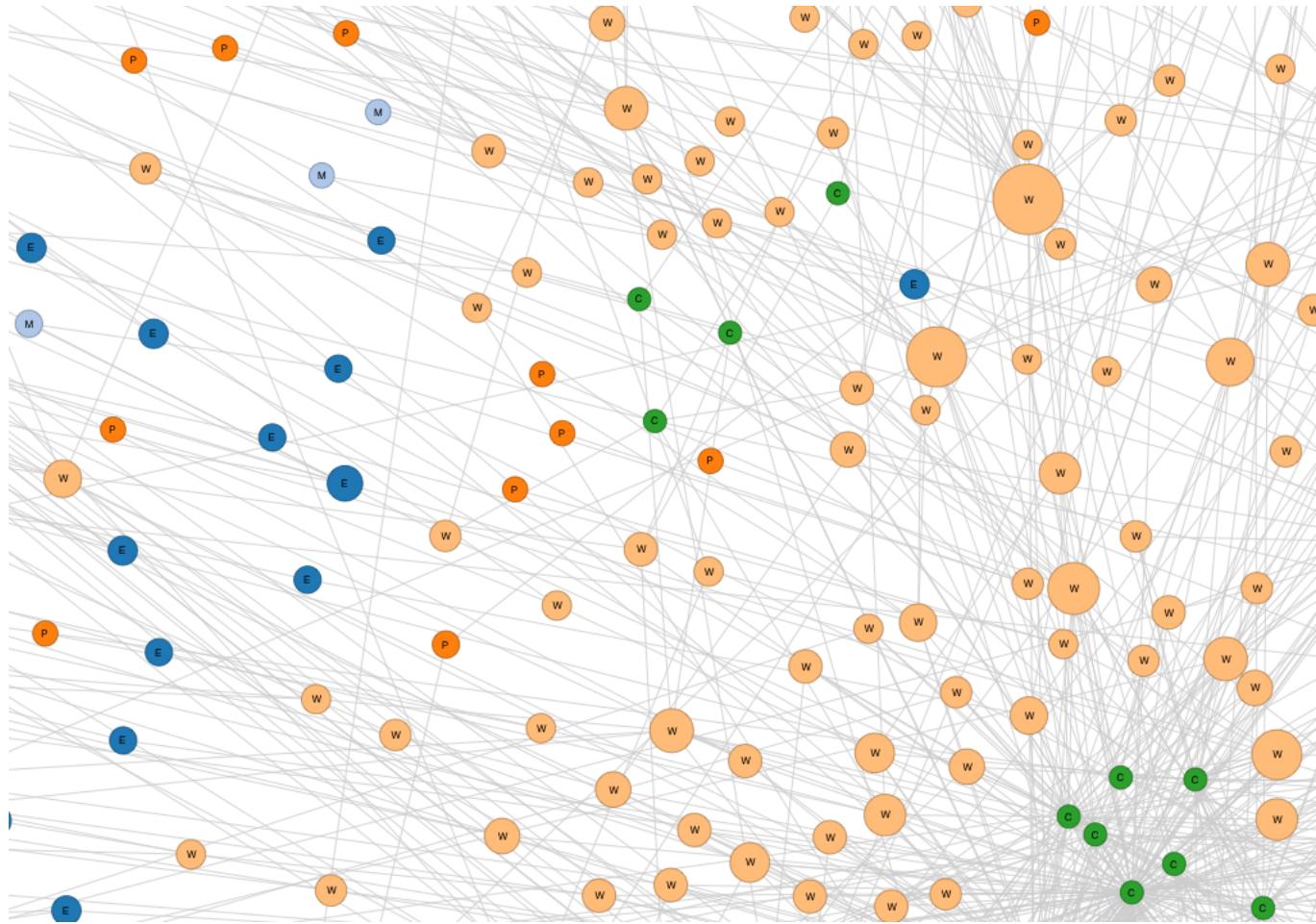


Fig. 66. Extract of a graph-based visualization of the FRBRized BIB-RCAT dataset in FRBR-ML

Figure 67 depicts (an extract of) the FRBRized collection of the BIB-RCAT dataset using Variations (Scherzo).

Basic Results

dijon.idi.ntnu.no:8085/scherzo/SearchResults.action?sKeyword=%3A&sManifType=&submit=Submit&_sourcePage=-uULSJPybNGdWlhUtBlb3hooItNjLffRTW... Naimdjoni

INDIANA UNIVERSITY

Scherzo

New search

Search: ":" as Keyword.

[Modify Search](#)

Browse Results By:

- Creator/Composer

[hide...](#)

Performer/Conductor or Arranger/Editor

Works: 207 results for ":" as Keyword

- La Fontaine, Jean de
[show...](#)
- Brunhoff, Jean de
[show...](#)
- Guillot, Ren
[show...](#)
- Daudet, Alphonse
[show...](#)
- Molière
[show...](#)

Recordings/Scores: 560 results for ":" as Keyword

- Fables. racontés par Louis de Funès; illustrations Bruno Robert. Paris Formulette production 17/06/2015.
Contents: Les fables de La Fontaine
People: La Fontaine, Jean de
Copies: [See Catalog](#)
- Histoire de Babar : le petit l'phant. Jean de Brunhoff. Vanves (Hauts-de-Seine) Hachette Jeunesse 09/06/2015.
Contents: Histoire de Babar
People: Brunhoff, Jean de
Copies: [See Catalog](#)
- Crin-Blanc. Ren Guillot; d'après le film d'Albert Lamorisse; illustrations Akos Szabo. Paris Le Livre de poche jeunesse 09/06/2015.
Contents: Crin-Blanc
People: Guillot, Ren
Copies: [See Catalog](#)
- Les lettres de mon moulin. Alphonse Daudet. Paris Librio 28/08/2013.
Contents: Lettres de mon moulin
People: Daudet, Alphonse
Copies: [See Catalog](#)
- Les fourberies de Scapin. Molière. Paris Librio 28/08/2013.
Contents: Les fourberies de Scapin
People: Molière
Copies: [See Catalog](#)
- La belle et la bête. Riquet la houppe. La Belle aux cheveux d'or La belle et la bête; Riquet la houppe; La Belle aux cheveux d'or. madame Leprince de Beaumont; édition présentée par Nicolas Saulais. Paris Nathan 23/08/2013.
Contents: La Belle et la Bte
People: Leprince de Beaumont, Jeanne-Marie
Copies: [See Catalog](#)
- pièce anonyme La farce de maître Pathelin. édition présentée, annotée et commentée par Thierry Revol; traduction de Alain Mié. Paris Larousse 21/08/2013.
Copies: [See Catalog](#)
- Les allumettes suédoises. Robert Sabatier; lu par Jean Barrier. La Bazoge (Sarthe) Cdl Ed. livres audio 28/06/2013.
Contents: Les allumettes suédoises
People: Sabatier, Robert
Copies: [See Catalog](#)
- Yvain ou Le chevalier au lion : extraits choisis. Chrétien de Troyes; traduction en français moderne et appareil pédagogique fablis par Hélène Dardelin; lexique fabli par Michèle Sendre. Paris Magnard 21/06/2013.
Contents: Yvain ou Le chevalier au lion
People: Chrétien de Troyes; Dardelin, Hélène

Fig. 67. Extract of the FRBRized BIB-RCAT dataset in Scherzo (Variations/VFRBR)

Below we present the interpretation of a MARC record in FRBR using XC. Figure 68 depicts the initial MARC record. Figures 69, 70 and 71 respectively illustrate the FRBRization into a Work, an Expression and a Manifestation.

Status: Active
Repository: MSTOA
Repository URL: <http://localhost:8080/OAIToolkit/oai-request.do>
Schema: marc21
Sets: MSTOA/Bibliographic records , MSTOA
Created at: 2016-01-08 09:58:29.0
Updated at: 2016-01-08 10:02:03.0
OAI datestamp:
OAI identifier: oai:sytrisresearch:datasetT42/11

3 Successors

```
<marc:record xmlns:marc="http://www.loc.gov/MARC21/slim" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.loc.gov/MARC21/slim http://localhost:8080/OAIToolkit/schema/MARC21slim_custom.xsd">
  <marc:leader>01208cam a2200277 i 4500</marc:leader>
  <marc:controlfield tag="005">08030518101543.0</marc:controlfield>
  <marc:controlfield tag="008">080730s1976 dcm f000 0 eng d</marc:controlfield>
  <marc:controlfield ind1=" " ind2=" " tag="020">
    <marc:subfield code="a">9782081209539</marc:subfield>
    <marc:subfield code="c">2,70 euro</marc:subfield>
  </marc:controlfield>
  <marc:controlfield ind1="1" ind2=" " tag="041">
    <marc:subfield code="a">French</marc:subfield>
  </marc:controlfield>
  <marc:controlfield ind1="1" ind2=" " tag="100">
    <marc:subfield code="a">Keime, Christian</marc:subfield>
    <marc:subfield code="4">drt</marc:subfield>
  </marc:controlfield>
  <marc:controlfield ind1="1" ind2="0" tag="240">
    <marc:subfield code="a">Roman de Renart</marc:subfield>
  </marc:controlfield>
  <marc:controlfield ind1="1" ind2="0" tag="245">
    <marc:subfield code="a">Roman de Renart (Le)</marc:subfield>
    <marc:subfield code="c">Christian Keime; Monique Lachet-Lagarde</marc:subfield>
    <marc:subfield code="h">[publication]</marc:subfield>
  </marc:controlfield>
  <marc:controlfield ind1="3" ind2=" " tag="260">
    <marc:subfield code="a">Paris</marc:subfield>
    <marc:subfield code="b">Flammarion</marc:subfield>
    <marc:subfield code="c">2008</marc:subfield>
  </marc:controlfield>
  <marc:controlfield ind1=" " ind2=" " tag="300">
    <marc:subfield code="a">159 pages</marc:subfield>
    <marc:subfield code="b">Illustrations, cover with colorful illustrations</marc:subfield>
    <marc:subfield code="c">18 cm</marc:subfield>
  </marc:controlfield>
  <marc:controlfield ind1="0" ind2=" " tag="490">
    <marc:subfield code="a">Etonnans classiques. 335</marc:subfield>
  </marc:controlfield>
  <marc:controlfield ind1="1" ind2=" " tag="700">
    <marc:subfield code="a">Lachet-Lagarde, Monique</marc:subfield>
    <marc:subfield code="4">adp</marc:subfield>
  </marc:controlfield>
  <marc:controlfield ind1="1" ind2=" " tag="760">
    <marc:subfield code="w">colauth-007</marc:subfield>
    <marc:subfield code="t">Etonnans classiques</marc:subfield>
  </marc:controlfield>
</marc:record>
```

MetadataServicesToolkit v 1.5.5 - svn.version:

Fig. 68. A MARC record displayed in XC web interface

Applications Places Fri Jan 8, 10:23 AM

Browse Records - Iceweasel

/manager | Browse Records | Search | dev

localhost:8080/MetadataServicesToolkit/st/viewRecord.action?recordId=52&query=&selectedFacetNames=[processed_from&selx C]

No process running

Metadata Services Toolkit

Hi admin | My Account | Logout

Repository | Harvest | Services | Processing Rules | Browse Records | Logs | Users/Groups | Configuration

Browse Records > Search Results > View Record oai:syrtis.research:MetadataServicesToolkit/marctoxctransformation/52

Status: Active
 Service: marctoxctransformation
 Schema: xc
 Sets:
 Created at: 2016-01-08 10:03:24.0
 Updated at: 2016-01-08 10:03:27.0
 OAI timestamp:
 OAI identifier: oai:syrtis.research:MetadataServicesToolkit/marctoxctransformation/52

1 Predecessor

```
<xc:frbr xmlns:xc="http://www.extensiblecatalog.info/Elements" xmlns:dcterms="http://purl.org/dc/terms/" xmlns:rdarole="http://rdvocab.info/roles" xmlns:rdvocab="http://rdvocab.info/Elements" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <xc:entity id="oai:syrtis.research:MetadataServicesToolkit/marctoxctransformation/52" type="work">
    <rdvocab:titleOfWork>Roman de Renart</rdvocab:titleOfWork>
  </xc:entity>
</xc:frbr>
```

MetadataServicesToolkit v 1.5.5 - svn.version:

Fig. 69. A FRBR Work entity displayed in XC web interface

Applications Places Fri Jan 8, 10:24 AM

Browse Records - Iceweasel

/manager | Browse Records | Search | dev

localhost:8080/MetadataServicesToolkit/st/viewRecord.action?recordId=53&query=&selectedFacetNames=[processed_from&selx C]

No process running

Metadata Services Toolkit

Hi admin | My Account | Logout

Repository | Harvest | Services | Processing Rules | Browse Records | Logs | Users/Groups | Configuration

Browse Records > Search Results > View Record oai:syrtis.research:MetadataServicesToolkit/marctoxctransformation/53

Status: Active
 Service: marctoxctransformation
 Schema: xc
 Sets:
 Created at: 2016-01-08 10:03:24.0
 Updated at: 2016-01-08 10:03:27.0
 OAI timestamp:
 OAI identifier: oai:syrtis.research:MetadataServicesToolkit/marctoxctransformation/53

1 Predecessor

```
<xc:frbr xmlns:xc="http://www.extensiblecatalog.info/Elements" xmlns:dcterms="http://purl.org/dc/terms/" xmlns:rdarole="http://rdvocab.info/roles" xmlns:rdvocab="http://rdvocab.info/Elements" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <xc:entity id="oai:syrtis.research:MetadataServicesToolkit/marctoxctransformation/53" type="expression">
    <rdarole:director>Keime, Christians</rdarole:director>
    <xc:titleOfExpression>Roman de Renart</xc:titleOfExpression>
    <rdvocab:illustrationContent>Illustrations, cover with colorful illustrations</rdvocab:illustrationContent>
    <xc:contributor>Lachet-Lagarde, Monique</xc:contributor>
    <xc:workExpressed>oai:syrtis.research:MetadataServicesToolkit/marctoxctransformation/52</xc:workExpressed>
  </xc:entity>
</xc:frbr>
```

MetadataServicesToolkit v 1.5.5 - svn.version:

Fig. 70. A FRBR Expression entity displayed in XC web interface

Applications Places Fri Jan 8, 10:22 AM

Browse Records - Iceweasel

/manager Browse Records Browse Records Browse Records Browse Records Browse Records Browse Records Browse Records

localhost:8080/MetadataServicesToolkit/st/viewRecord.action?recordId=51&query=&selectedFacetNames=[processed_from&selv C]

Search

Metadata Services Toolkit

Hi admin | My Account | Logout

Repository Harvest Services Processing Rules Browse Records Logs Users/Groups Configuration

Browse Records Search Results View Record oai:syrtis.research:MetadataServicesToolkit/marctoxctransformation/51

Status: Active
 Service: marctoxctransformation
 Schema: xc
 Sets:
 Created at: 2016-01-08 10:03:24.0
 Updated at: 2016-01-08 10:03:27.0
 OAI timestamp:
 OAI identifier: oai:syrtis.research:MetadataServicesToolkit/marctoxctransformation/51

1 Predecessor

```

<xc:frbr xmlns:xc="http://www.extensiblecatalog.info/Elements" xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:rdvocab="http://rdvocab.info/roles" xmlns:rdvocab="http://rdvocab.info/Elements"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <xc:entity id="oai:syrtis.research:MetadataServicesToolkit/marctoxctransformation/51" type="manifestation">
    <rdvocab:statementOfResponsibilityRelatingToTitle>Christian Keime; Monique Lachet-
    Lagarde</rdvocab:statementOfResponsibilityRelatingToTitle>
    <dcterms:title>Roman de Renart (Le)</dcterms:title>
    <rdvocab:placeOfProduction>Paris</rdvocab:placeOfProduction>
    <dcterms:publisher>Flammarion</dcterms:publisher>
    <dcterms:issued>2008</dcterms:issued>
    <dcterms:extent>159 pages</dcterms:extent>
    <rdvocab:dimensions>18 cm</rdvocab:dimensions>
    <dcterms:isPartOf>Etonnantes classiques. 335</dcterms:isPartOf>
    <dcterms:isPartOf>Etonnantes classiques</dcterms:isPartOf>
    <xc:expressionManifested>oai:syrtis.research:MetadataServicesToolkit/marctoxctransformation
    /53</xc:expressionManifested>
  </xc:entity>
</xc:frbr>
```

MetadataServicesToolkit v 1.5.5 - svn.version:

Fig. 71. A FRBR Manifestation entity displayed in XC web interface

7 Acknowledgments

This work has been partially supported by the French Agency ANRT (www.anrt.asso.fr), the company PROGILONE (www.progilone.com/), a PHC Aurora funding (#34047VH) and a CNRS PICS funding (#PICS06945).

References

1. Riva, P.: Mapping MARC 21 Linking Entry Fields to FRBR and Tilletts Taxonomy of Bibliographic Relationships. *Library resources & technical services* 48(2), 130–143 (2013)
2. Zhong, N., Li, Y., Wu, S.T.: Effective pattern discovery for text mining. *Transactions on Knowledge and Data Engineering* 24(1), 30–44 (2012)