

A Heuristic Method for Large-Scale Cognitive-Diagnostic Computerized Adaptive Testing

Jill-Jênn Vie

RIKEN Center for Advanced
Intelligence Project (AIP)
Tokyo, Japan
vie@jill-jenn.net

Fabrice Popineau

Laboratoire de recherche
en informatique (LRI)
Orsay, France
fabrice.popineau@lri.fr

Françoise Tort

ENS Paris-Saclay
Cachan, France
francoise.tort
@stef.ens-cachan.fr

Benjamin Marteau

French Ministry of Education
Paris, France
benjamin.marteau
@education.gouv.fr

Nathalie Denos

Université Grenoble-Alpes
Grenoble, France
nathalie.denos
@univ-grenoble-alpes.fr

Context: Cognitive-Diagnostic Computerized Adaptive Testing

Input: a dependency graph between **knowledge components** (KC)

Output: an adaptive test (CAT tree) with feedback

- **Adaptive testing** \Rightarrow personalized assessment
- **Formative testing** \Rightarrow need of feedback (using the KCs)
- **Many knowledge components** \Rightarrow standard methods do not apply
- **Cold-start** \Rightarrow no user data available at the time

Item Response Theory

Students $i \in I$ have **unknown** level θ_i
Questions $j \in J$ have difficulty d_j (potentially calibrated on data)

$$\Pr(\text{correct}_{ij}) \triangleq \Pr(\text{student}_i \text{ answers correctly question}_j) \triangleq \frac{1}{1 + e^{-(\theta_i - d_j)}}$$

$\Rightarrow \theta_i$ is easily estimated (maximum likelihood), but no feedback

DINA & Attribute Hierarchy Model

Knowledge components $1, \dots, K$
Students $i \in I$ have **unknown** knowledge $\in \{0, 1\}^K$
Questions $j \in J$ have requirements $\in \{0, 1\}^K$ (q-matrix), slip s_j and guess g_j parameters

$$\Pr(\text{correct}_{ij}) \triangleq \begin{cases} 1 - s_j & \text{if student}_i \text{ masters every requirement of question}_j \\ g_j & \text{otherwise.} \end{cases}$$

\Rightarrow can be used in cold-start, but not scalable when K grows (≥ 50)

Can we do better?

Our Heuristic Method

Knowledge components $1, \dots, K$ contain a **tag** and a **difficulty** level
are linked in a **dependency graph**
Students $i \in I$ have **unknown** level θ_i and knowledge $\in \{0, 1\}^K$
Questions $j \in J$ have a unique main requirement $\in \{1, \dots, K\}$

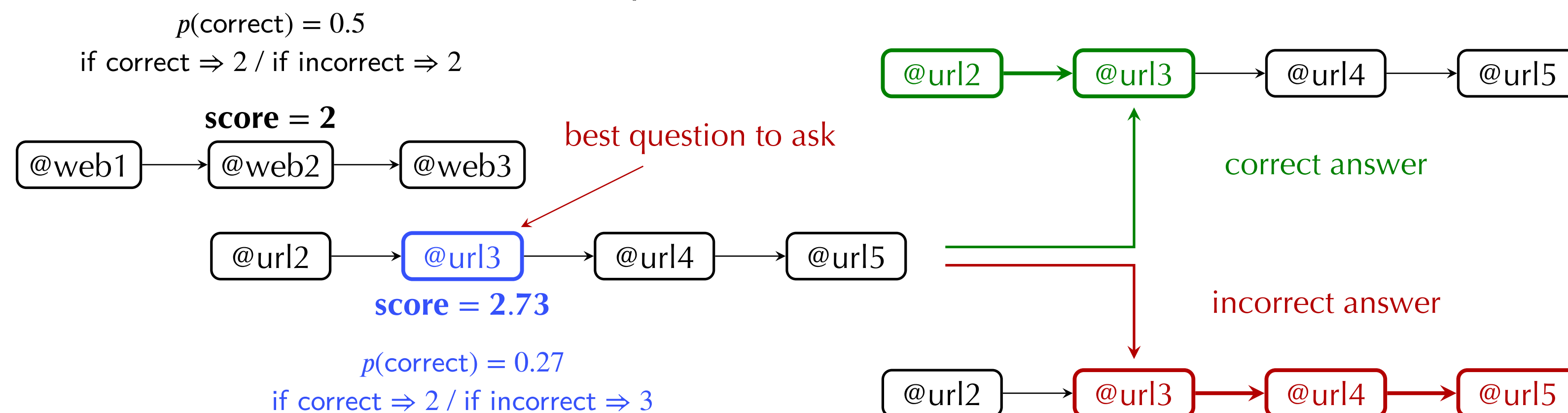
$$\Pr(\text{correct}_{ij}) \text{ is the same as in Item Response Theory}$$

$$\text{score}(j) = \Pr(\text{correct}_{ij}) \cdot N_{\text{acquired nodes if correct}} + \Pr(\text{incorrect}_{ij}) \cdot N_{\text{non-acquired nodes if incorrect}}$$

- **Rough estimate of the number of nodes marked** in the dependency graph at each step
- **Greedy selection** to build the tree
- Later in the test, we can **switch to a more precise diagnostic model**
- **Add edges to dependency graph** based on usage (Deep Knowledge Tracing, NIPS 2015)

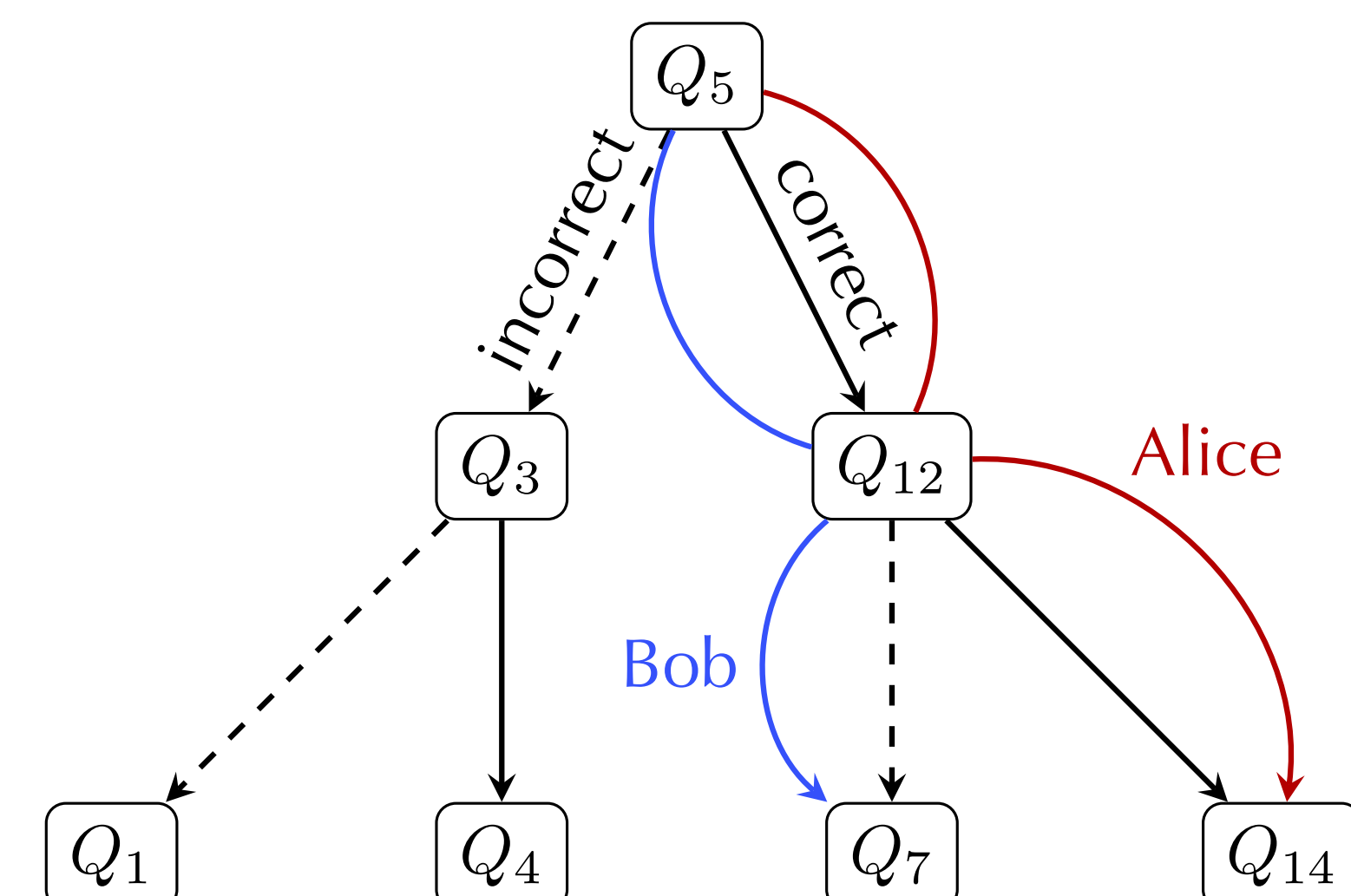
Greedy Selection

If the estimated level of the student $\hat{\theta}_i$ is 2:



Computerized Adaptive Testing (CAT)

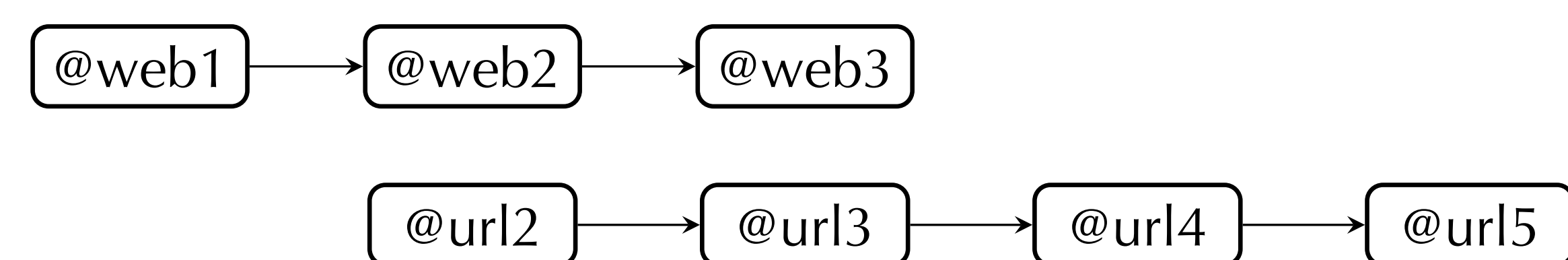
The next question is chosen based on the answer history



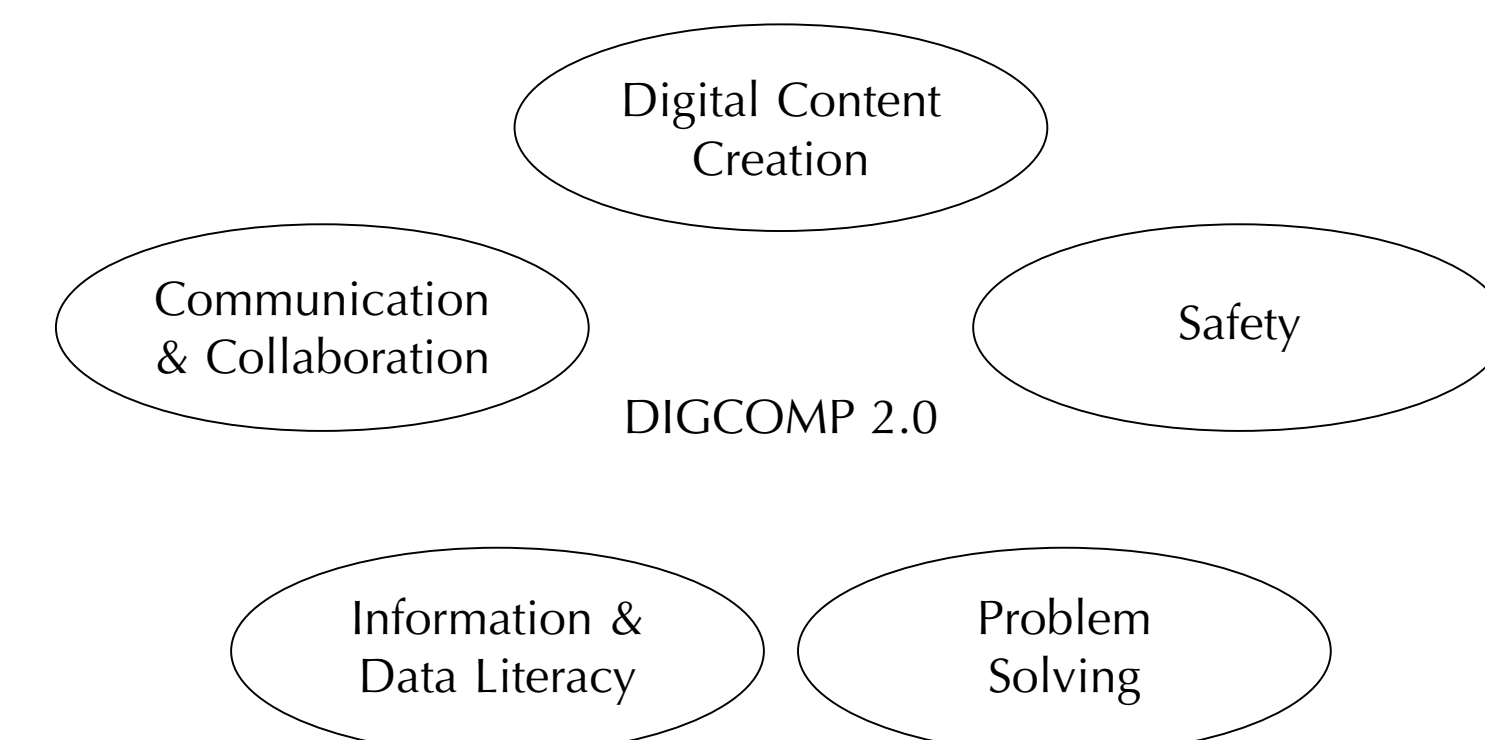
Input: Dependency Graph over DIGCOMP 2.0

Nodes: **800 knowledge components** of difficulty 1–5

Edges are prerequisites ($u \rightarrow v$ means, u should be mastered before v)



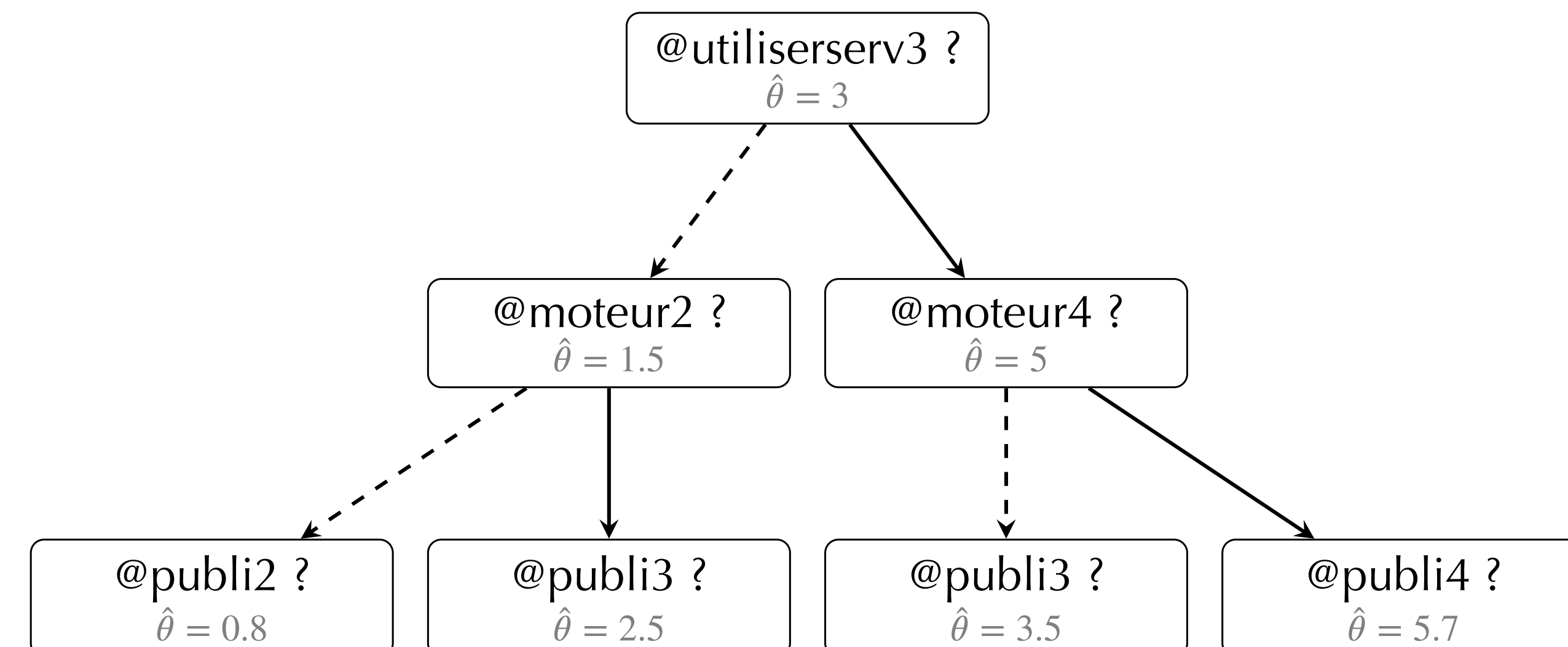
Based on DIGCOMP 2.0, the European Digital Competence Framework



Each question is mapped to a unique (required) knowledge component

Consigne	acquis
Écrivez en B2 une formule qui compte le nombre de lettres de B1....	@calcul5
En utilisant les filtres des colonnes, trouvez le nombre de personnes de ...	@extraire2
Le fichier contient un tableau présentant les livres les plus prêtés à Paris ...	@extraire3
N'affichez que les personnes de Calais ayant 20 ans et une note ...	@extraire4

Output: CAT tree



Application: Certifying the Digital Competencies of French citizens

- **800 skeletons of exercises** based on **evidence-centered design**
- **16 non-adaptive tests** measuring **50 knowledge components** over **5 difficulties**
- Output: **3 adaptive tests** of **20 questions**
- So far, **514731 collected answers** in the database from **36953 users**

High school user data will be collected from September 2017:

- **3.5 million** high school students (grade 8 to 12)
- **1.25 million** higher-education students

The platform is on GitHub (license AGPLv3)

