

Proyecto FONDEF:

Desarrollo y Validación de un Sistema Automatizado de Revisión de Ítems (SARI)

Concurso IDeA I+D 2021



PLAN DE LA PRESENTACIÓN

- 1. En resumen, ¿qué es SARI?**
- 2. ¿Cómo se ha construido y qué hace?**
- 3. Demo y Historial**
- 4. Logros, Aprendizajes, Proyecciones**
- 5. Preguntas**

Problema y Solución Propuesta

La calidad de las preguntas de selección múltiple es un tema central en educación. Estas preguntas se usan muy frecuentemente para evaluar el aprendizaje, toma mucho tiempo construirlas, e incluso cuando provienen de pruebas estandarizadas, a veces presentan imperfecciones de construcción, que pueden afectar negativamente la medición.

Desarrollamos un **Sistema Automático de Revisión de Ítems (SARI)** que detecta automáticamente 24 tipos de imperfecciones. Fue transferido a nuestras 2 entidades asociadas.



Cronograma

*Inicio del proyecto: **diciembre 2021***

AÑO 1: Desarrollo y Validación de algoritmos de detección

Entrenamiento

Incorporación en SARI y bancos de ítems



AÑO 2: Estudios de calidad del sistema SARI

Estudio de eficiencia

Estudio de utilidad



PRODUCTO FINAL: SARI prototipo de servicio de revisión automatizada de ítems

Perfeccionamiento ergónomico

Funcional para PAES y DIA

*Final del proyecto: **noviembre 2023***

¿Qué es? Sari

Una plataforma web

en un servidor local

con datos y acceso securizados

Un sistema de comunicación

con un editor online y procesador de textos

y revisiones instantáneas

¿Qué hace SARI?

1. **Etiqueta** las imperfecciones de construcción detectadas
2. **Señala**, cuando le es posible, qué parte del texto podría requerir edición
3. **Entrega un consejo** general sobre el problema asociado a la imperfección
4. **Permite chequear** que la versión editada está ahora libre de imperfecciones

¿Qué no hace SARI?

1. No entrega recomendación específica de modificación
2. No elimina automáticamente los problemas señalados
3. NO APLICA MODIFICACIÓN AL ÍTEM

Algoritmos

Equipo



Abelino Jiménez

PhD in Electrical and
Computer Engineering,
CMU, USA

Master in Electrical and
Computer Engineering,
CMU, USA

Ingeniero Civil
Matemático, Universidad
de Chile



Matías Altamirano

PhD student in Statistics,
UCL, UK

Magíster en
Matemáticas Aplicadas,
Universidad de Chile

Ingeniero Civil
Matemático, Universidad
de Chile



Danner Schlotterbeck

Magíster en
Matemáticas Aplicadas,
Universidad de Chile

Ingeniero Civil
Matemático, Universidad
de Chile



Diego Reyes

Magíster en Data
Science, Universidad de
Chile

Ingeniero Civil Industrial,
Universidad de Chile



Objetivos de SARI como sistema

Detectar las imperfecciones de cada ítem,
proponer una solución para eliminarlas,
y **verificar** versiones mejoradas del ítem

Nuestro objetivo se enfoca en **Detectar imperfecciones** y **Proponer Soluciones**.

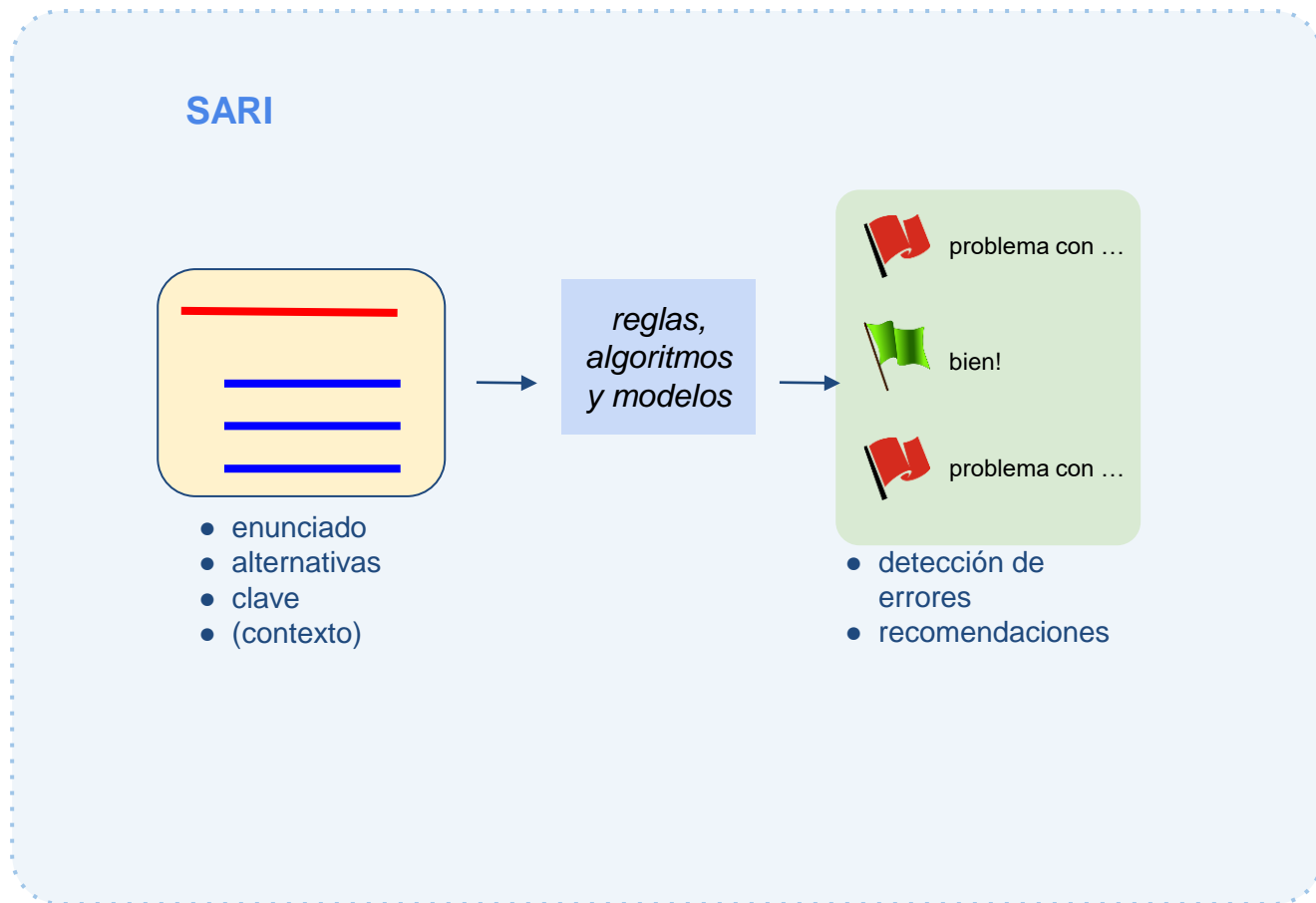
La lista de imperfecciones...

N°	Nivel	IMPERFECCIÓN	DETECCION		
		Regla no respetada (de Haladyna et al., 2002, traducido por Moreno et al., 2004):	Estrategia a usar	Fuente	Nuestra predicción de eficacia
1	CONTENIDO	Cada ítem debería reflejar un contenido específico y una única conducta mental específica, tal como sea requerido en las especificaciones del test (tabla de doble entrada, proyecto del test).	requiere información adicional	idea propia	SI
2		Base cada ítem en un contenido importante para el aprendizaje; evite contenidos triviales.	análisis semántico	idea propia	EN PARTE
3		Use material novedoso para evaluar el aprendizaje de alto nivel. Cuando los utilice en un ítem, parafrasee el lenguaje de los libros de texto, o el lenguaje utilizado durante la instrucción, para así evitar evaluar el mero recuerdo.	análisis semántico	idea propia	EN PARTE
4		Mantenga el contenido de cada ítem independiente del contenido de otros ítems del test.	análisis de ensamblaje	idea propia	SI
5		Evite contenidos muy específicos o muy generales.	por definir	no hay	NO
6		Evite ítems basados en opiniones.	análisis semántico	idea propia	SI
7		Evite ítems con trampas.	análisis comparativo entre coincidencia tallo-respuesta correcta y tallo-distractores	García 2011	EN PARTE
8		Use un vocabulario sencillo para el grupo de estudiantes que están siendo evaluados.	análisis semántico	idea propia	SI
9	FORMATO	Del formato convencional de elección múltiple utilice la interrogación, completar frases, la mejor respuesta, elección alternativa, verdadero-falso, verdadero-falso múltiple, emparejamiento, los conjuntos de ítems y los dependientes de contexto; sin embargo, evite el formato de elección múltiple complejo (el tipo K).	análisis presencia ciertas palabras o caracteres	idea propia	SI
10	ESTILO	Construya el ítem de forma vertical, no horizontal.	análisis visual	idea propia	SI
11		Corrija y pruebe los ítems.	análisis semántico y sintáctico	idea propia	SI
12		Utilice una gramática, puntuación, mayúsculas y minúsculas y deletreo correctos.	análisis semántico y sintáctico	idea propia	SI
13	TALLO (encabezado)	Minimice la cantidad de lectura en cada ítem.	análisis de extensión de encabezado	García 2011	SI
14		Asegúrese de que el sentido del enunciado resulta muy claro.	análisis de extensión de encabezado	Brunnquell 2011	EN PARTE
15		Incluya la idea central en el enunciado y no en las opciones.	análisis de extensión de opciones	Brunnquell 2011 + García 2011	EN PARTE
16		Evite adornar el texto en exceso (palabrería excesiva).	análisis de extensión de encabezado	García 2011	EN PARTE
17		Expresé el enunciado de manera afirmativa, evitando términos negativos tales como NO o EXCEPTO. Si usa términos negativos, hágalo con sumo cuidado y asegúrese que aparecen en mayúsculas o negritas.	análisis presencia ciertas palabras o caracteres	Brunnquell 2011 + García 2011	SI
18		Escriba tantas opciones como pueda, si bien la investigación sugiere que con tres es suficiente.	análisis presencia ciertas palabras o caracteres	idea propia + Mkrtychyan 2011	EN PARTE
19		Asegúrese que sólo una de esas opciones es la respuesta correcta.	por definir	no hay	NO
20	OPCIONES	Varié la colocación de la respuesta correcta según el número de opciones.	análisis de distribución de opciones en prueba ensamblada	idea propia	SI
21		Coloque las opciones en un orden lógico o numérico.	análisis presencia ciertas palabras o caracteres	idea propia	EN PARTE
22		Construya las opciones independientes entre sí, no deben solaparse.	análisis presencia ciertas palabras o caracteres, comparativo entre opciones	García 2011	EN PARTE
23		Mantenga a las opciones homogéneas en contenido y estructura gramatical.	análisis semántico y sintáctico	García 2011 + Pho 2015	SI
24		Escriba las opciones con una longitud aproximadamente igual.	análisis de extensión de opciones	Brunnquell 2011 + García 2011	SI
25		La opción Ninguna de las anteriores debe usarse con mucho cuidado.	análisis presencia ciertas palabras o caracteres	idea propia	SI
26		Evite la opción Todas las anteriores.	análisis presencia ciertas palabras o caracteres	idea propia	SI
27		Escriba las opciones de forma afirmativa, evite términos negativos tales como NO.	análisis presencia ciertas palabras o caracteres	García 2011	SI
28		Evite dar pistas sobre la respuesta correcta, tales como: Determinantes específicos como siempre, nunca, completamente y absolutamente.	análisis presencia ciertas palabras o caracteres	Brunnquell 2011 + García 2011	SI
29		Evite dar pistas sobre la respuesta correcta, tales como: Asociaciones por sonido similar y opciones idénticas o parecidas a términos del enunciado.	análisis presencia ciertas palabras o caracteres	Brunnquell 2011 + García 2011	EN PARTE
30		Evite dar pistas sobre la respuesta correcta, tales como: Inconsistencias gramaticales que indiquen al sujeto la elección correcta.	análisis semántico y sintáctico	idea propia	SI
31		Evite dar pistas sobre la respuesta correcta, tales como: Opción correcta destacada.	análisis de extensión de opciones	García 2011	EN PARTE
32		Evite dar pistas sobre la respuesta correcta, tales como: Pares o trios de opciones que indiquen al sujeto la opción correcta.	análisis de repeticiones	Brunnquell 2011	EN PARTE
33		Evite dar pistas sobre la respuesta correcta, tales como: Opciones claramente absurdas o ridículas.	análisis semántico	idea propia	EN PARTE
34		Haga plausibles todos los distractores.	análisis semántico y sintáctico	idea propia + Mkrtychyan 2011	EN PARTE
35		Use errores usuales de los estudiantes para escribir los distractores.	por definir	no hay	NO
36	Use el humor si es compatible con el profesor y con el ambiente de aprendizaje.	por definir	no hay	NO	

Table 3
Downing and Haladyna's Model of Qualitative Item Validity Evidence

Type of Evidence	Activity	Evidence Needed
Content definition	Role delineation, job/task analysis, practice analysis	Documentation of the method(s) used to select item content
Test specifications	Table of specifications or test blueprint	Documentation of systematic link of test content to specifications or blueprint
Item writer training	Develop training materials and methods; train item writers	Documentation of methods, principles, written materials, and sample items
Adherence to item writing principles	Standard item writing rules adopted	Evidence of compliance with rules and documentation of process used to review items
Cognitive behavior	Cognitive classification system used to classify items	Documentation of system used and its rationale; reports of any research using system
Item content verification	Content experts review and judge items	Content experts' credentials; records of content expert review process
Item editing	Review items and professionally edit	Credentials and experience of editors; editorial and style guidelines
Bias-sensitivity review	Policies and procedures developed	Documentation of bias-sensitivity review; rationale for policies; credentials of reviewers
Item tryout and testing	Pre-test, pilot or field test items; item performance data; examinee interviews	Documentation of examinee pilot test data; examinee and item characteristics
Key validation and verification	Correctness of key answer verified by panel of content experts	Policy and procedures for key verification; documentation of key validation results
Test security plan	A test security policy and set procedures are developed	Copy of policy and procedures manual that specifies how items are protected from security lapses (Downing & Haladyna, 1997, p. 64)

SARI, como sistema, es un conjunto de reglas que analizan un ítem



La implementación de reglas de detección de imperfecciones contó con dos importantes desafíos



Determinar qué reglas son de alto impacto y factibles de implementar



Falta de Datos Anotados al comenzar el proyecto

Comenzamos con un enfoque tradicional de implementación



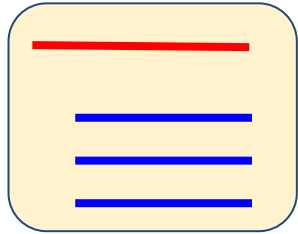
Implementación directa de algoritmos

- No basadas en datos
- En general, imperfecciones simples de detectar (Marcadores, regex, etc)

Formato	Evite el formato de elección múltiple complejo (el tipo K).
Enunciado	Expresé el enunciado de manera afirmativa.
	Incluya la idea central en el enunciado y no en las opciones. Enunciado cierra con pregunta explícita
Opciones	Evite la opción Todas las anteriores.
	La opción "Ninguna de las anteriores" debe usarse con mucho cuidado.
Pistas en Opciones	Evite dar pistas sobre la respuesta correcta, tales como la respuesta correcta es la más larga.
	Evite dar pistas sobre la respuesta correcta, tales como: Determinantes específicos como siempre, nunca, completamente y absolutamente.
Orden en Opciones	Coloque en orden las opciones si son numéricas
	Coloque en orden las opciones si tiene un solo término
	Coloque en orden las opciones de acuerdo a su extensión

Esto además nos permitió comenzar con una buena estructura del proyecto a nivel de software, en especial en términos de definiciones críticas

Desde el inicio del proyecto quedó determinado los campos que los algoritmos debían trabajar



- enunciado
- alternativas
- clave
- (contexto)



*reglas,
algoritmos y
modelos*



problema con ...
bien!
problema con ...

- detección de errores
- recomendaciones

```
item = SimpleItem(prompt,  
                  choices,  
                  key)
```



*reglas,
algoritmos y
modelos*



```
rule_results =  
analyzer.check(item)
```

```
prompt:      str  
  choices:  List[str]  
  key:      int
```

List[RuleResult]

```
RuleResult  
  ruleID: str  
  ruleDescription: str  
  ruleVersion: str  
  flag: bool  
  score: float  
  highlight: str  
  recommendation: str
```

Resultados
de la regla



Para implementar reglas más complicadas consideramos una aproximación desde el aprendizaje automático

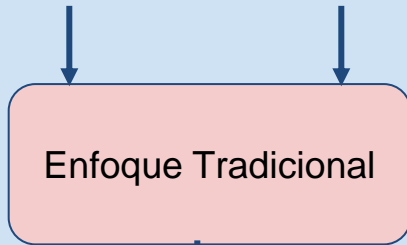


Entrenamiento de Modelos de Machine Learning

- Uso de datos entregados y anotaciones
- Asociado a reglas más difíciles de implementar

Regla

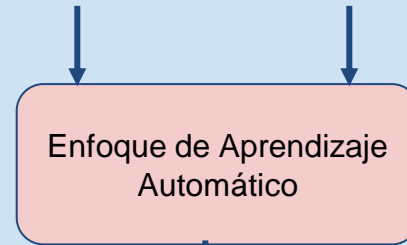
Data



Respuesta

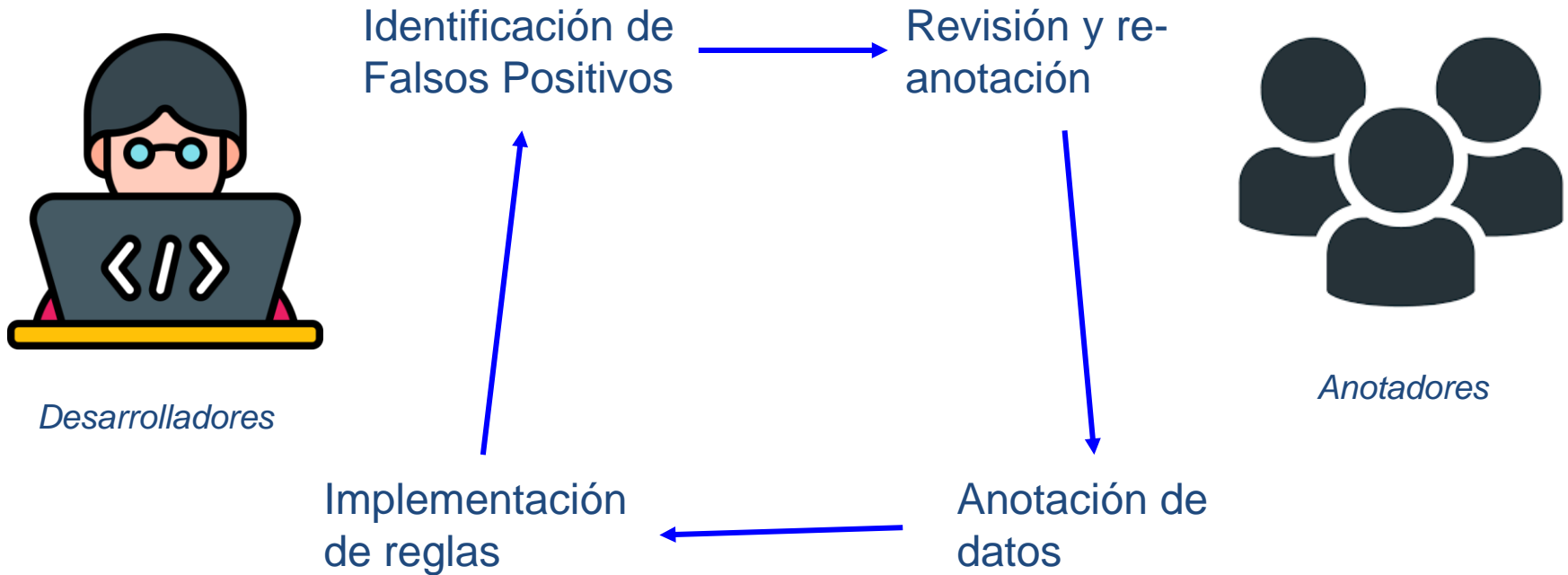
Respuesta

Data



Regla

En la práctica, el proceso de implementación de reglas fue un proceso iterativo entre desarrolladores y anotadores



WebApp para compartir implementación de SARI

Inserte pregunta

Ante un accidente, ¿cuál de las siguientes afirmaciones no es cierta?

- a) Se recomienda reconocerlo en el mismo lugar en que se encuentra.
- b) No se aconseja moverlo ni trasladarlo hasta que no se le hayan hecho la primera valoración de emergencia.
- c) Traslado puede agravar la situación o causarle nuevas lesiones.
- d) Lo primero es trasladarlo a un lugar seguro.

SOLUCION: d

Analizar!

Ante un accidente, ¿cuál de las siguientes afirmaciones no es cierta?

- ✗ a) Se recomienda reconocerlo en el mismo lugar en que se encuentra.
- ✗ b) No se aconseja moverlo ni trasladarlo hasta que no se le hayan hecho la primera valoración de emergencia.
- ✗ c) Traslado puede agravar la situación o causarle nuevas lesiones.
- ✓ d) Lo primero es trasladarlo a un lugar seguro.

► Regla 24

Descripción: **Escriba las opciones con una longitud aproximadamente igual.**

Score: 1.0

Recomendación: *La opción b) tiene muchos caracteres.*

Ante un accidente, ¿cuál de las siguientes afirmaciones no es cierta?

- a) Se recomienda reconocerlo en el mismo lugar en que se encuentra.
- b) **No se aconseja moverlo ni trasladarlo hasta que no se le hayan hecho la primera valoración de emergencia.**
- c) Traslado puede agravar la situación o causarle nuevas lesiones.
- d) Lo primero es trasladarlo a un lugar seguro.

SOLUCIÓN: d



En términos de resultados

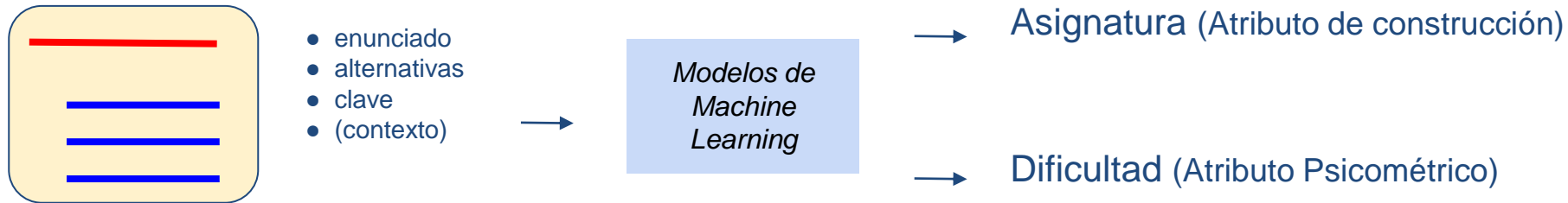
Precisión fue nuestra medida de desempeño.

Tasa de acierto cuándo regla declara que existe imperfección.

- 24 reglas implementadas
- 18 reglas con precisión mayor a 75%

Parte	Regla	Tipo de implementación	Precisión (mejoramiento)
Contenido	Cada ítem debería reflejar un contenido específico y una única conducta mental específica, tal como sea requerido en las especificaciones del test (tabla de doble entrada, proyecto del test).	Modelo Entrenado	<75%
Formato	Evite el formato de elección múltiple complejo (el tipo K).	Explícita	96%
Estilo	Utilice una gramática correcta.	Modelo Externo	-
	Utilice buena ortografía.	Modelo Externo	-
	Minimice la cantidad de lectura en cada ítem.	Modelo Entrenado	<75%
Enunciado	Expresar el enunciado de manera afirmativa.	Explícita	87%
	Incluya la idea central en el enunciado y no en las opciones.	Explícita	87%
	Asegúrese de que el sentido del enunciado resulta muy claro.	Modelo Entrenado	<75%
	Enunciado cierra con pregunta explícita	Explícita	100%
Opciones	Evite la opción Todas las anteriores.	Explícita	100%
	Escriba las opciones de forma afirmativa, evite términos negativos tales como NO.	Modelo Externo	91%
	La opción "Ninguna de las anteriores" debe usarse con mucho cuidado.	Explícita	100%
	Escriba las opciones con una longitud aproximadamente igual.	Model Entrenado	92%
	Construya las opciones independientes entre sí, no deben solaparse.	Model Entrenado	<75%
	Haga plausibles todos los distractores.	Model Entrenado	75%
	Asegúrese que sólo una de esas opciones es la respuesta correcta.	Model Entrenado	<75%
Pistas en Opciones	Evite dar pistas sobre la respuesta correcta, tales como la respuesta correcta es la más larga.	Explícita	78%
	Evite dar pistas sobre la respuesta correcta, tales como: Pares o tríos de opciones que indiquen al sujeto la opción correcta.	Model Entrenado	<75%
	Evite dar pistas sobre la respuesta correcta, tales como: Asociaciones por sonido similar y opciones idénticas o parecidas a términos del enunciado.	Modelo Externo	79%
	Evite dar pistas sobre la respuesta correcta, tales como: Determinantes específicos como siempre, nunca, completamente y absolutamente.	Explícita	78%
Orden en Opciones	Coloque en orden las opciones si son numéricas	Explícita	100%
	Coloque en orden las opciones si tiene un solo término	Explícita	100%
	Coloque en orden las opciones de acuerdo a su extensión	Explícita	100%
	Coloque en orden las opciones de acuerdo a su significado	Modelo Externo	-

Adicionalmente, durante este proyecto pudimos explorar métodos para inferir atributos de los ítems



Estas estimaciones contribuyen a la gestión de grandes volúmenes de datos de ítems, así como al mejoramiento de las preguntas para pilotaje.

Publicación de artículo en conferencia internacional



Methodologies and Intelligent Systems for
Technology Enhanced Learning

***"Multiple-choice Questions Difficulty
Prediction with Neural Networks"***

by Diego Reyes Troncoso, Abelino
Jimenez, Pablo Dartnell, and Séverin Lions



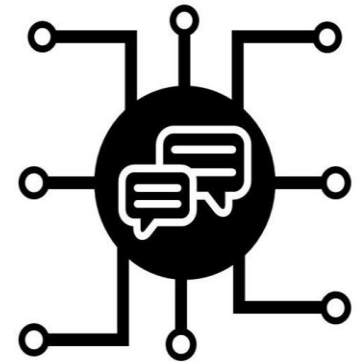
Oportunidades para el futuro



Procesamiento de imágenes en ítems



Mejoramiento de
compresión de Contenido
Matemático












Uso de LLMs









Demo (2022)

Bienvenido a SARI-APP!

Sistema Automatizado de Revision de Items

sari v0.0.17 - sari_app v0.0.11

B I U  x_2 x^2  **A**      H₁ H₂ Normal  Normal 

       Sans Serif 

Ante un accidente, ¿cuál de las siguientes afirmaciones no es cierta?

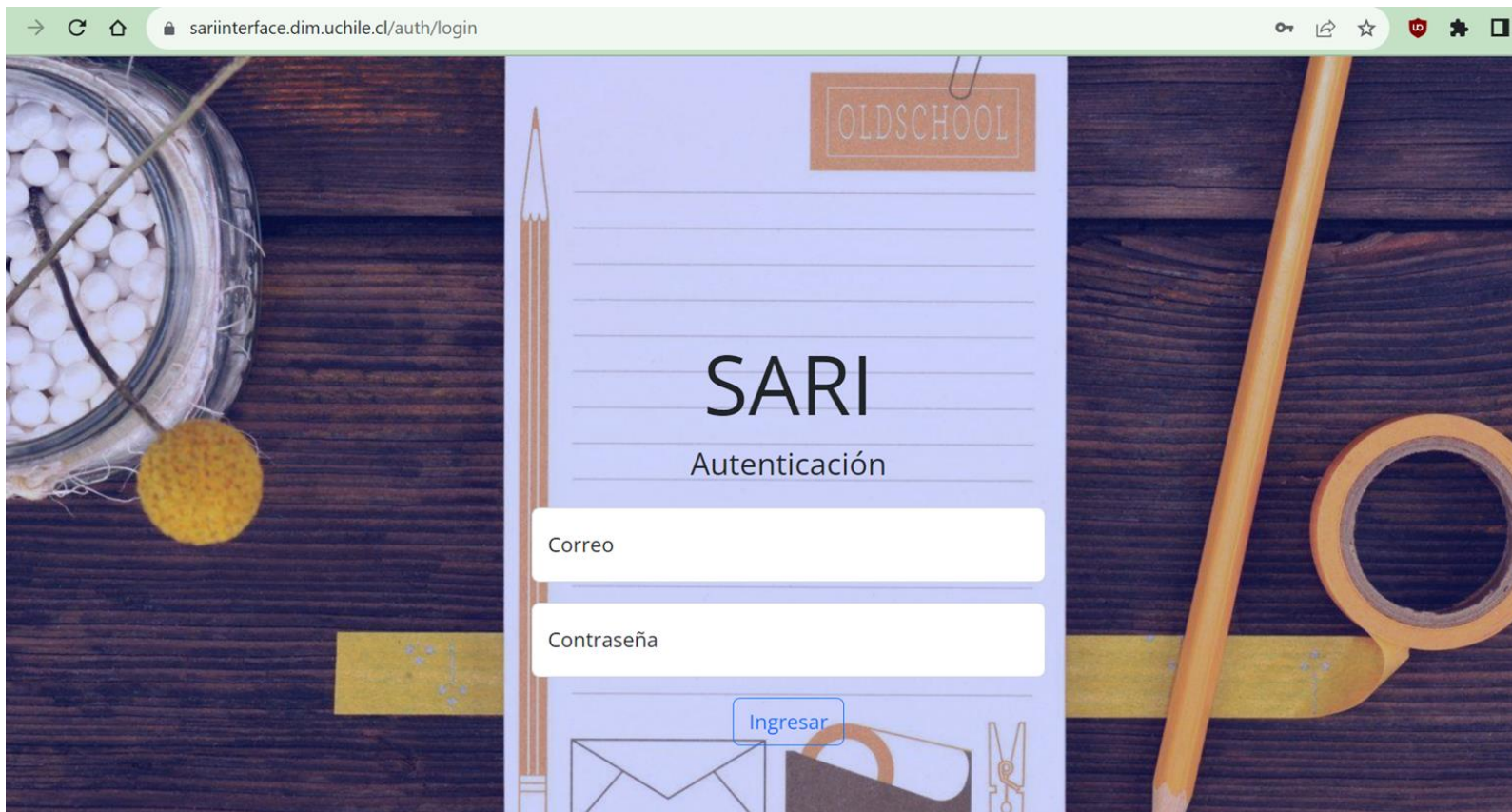
- a) Se recomienda reconocerlo en el mismo lugar en que se encuentra.
- b) No se aconseja moverlo ni trasladarlo hasta que no se le hayan hecho la primera valoración de emergencia.
- c) Traslado puede agravar la situación o causarle nuevas lesiones.
- d) Lo primero es trasladarlo a un lugar seguro.

SOLUCION: d

Analizar!



Portada (2023)



Demo (2024)

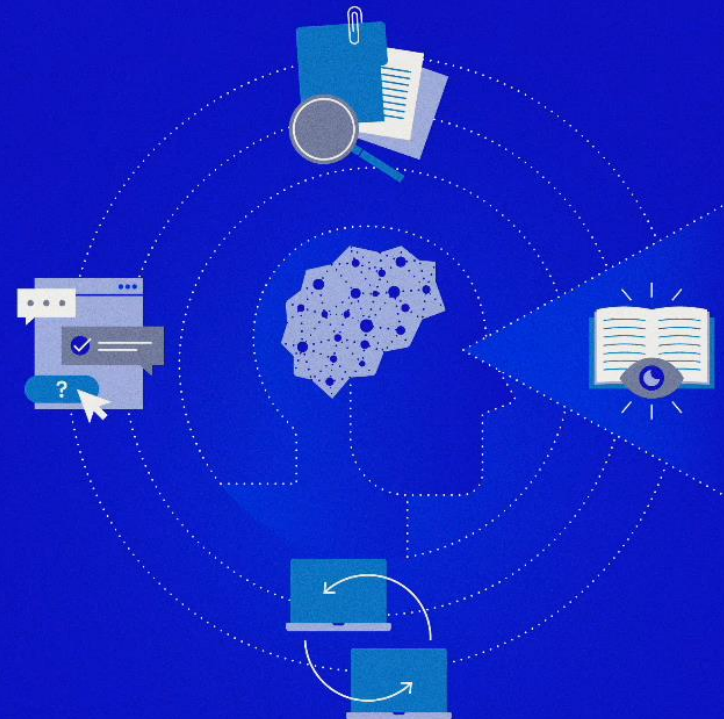


¡Bienvenido(a) a SARI!

Correo

Contraseña

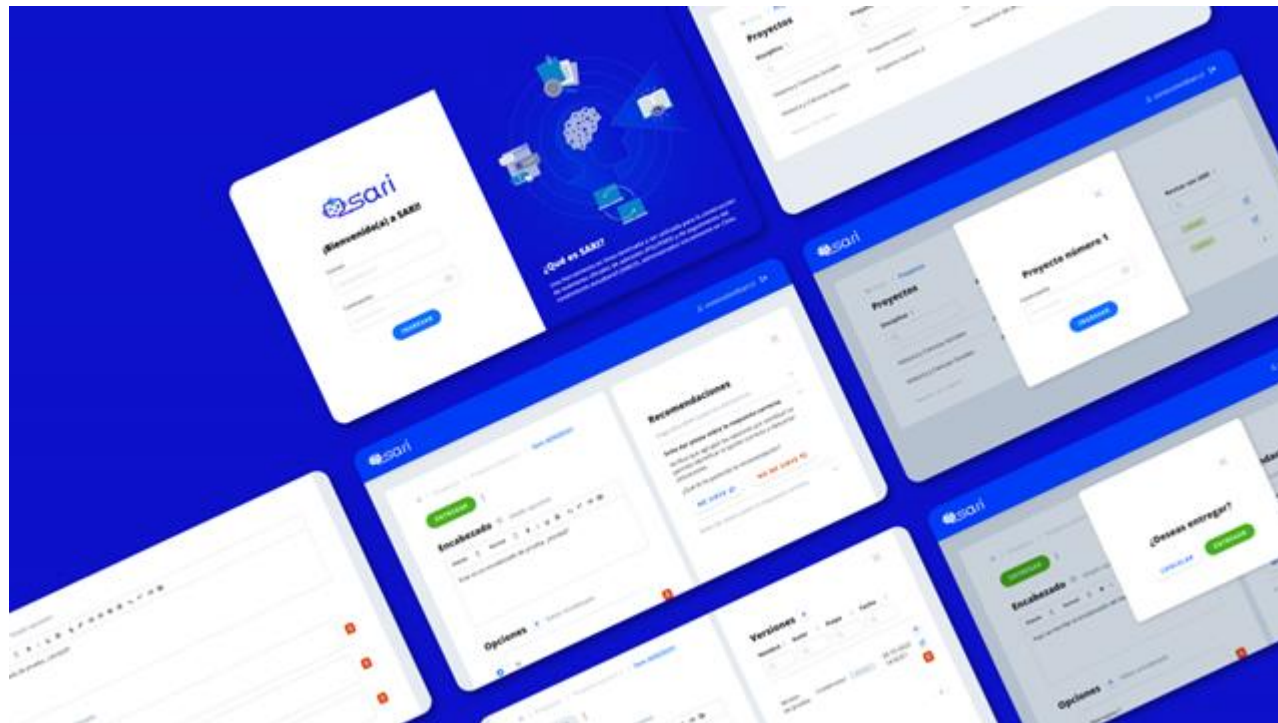
INGRESAR



¿Qué es SARI?

SARI es una herramienta en línea destinada a apoyar la construcción y revisión de los ítems que se incorporarán en las pruebas chilenas de más alta importancia: las PAES, las pruebas SIMCE, y las pruebas DIA.

Logros, Aprendizajes, Proyecciones



Anotación de ítems: 5471 ítems co-anotados, 32 criterios



Verónica Muñoz

**MATE-
MÁTICAS**

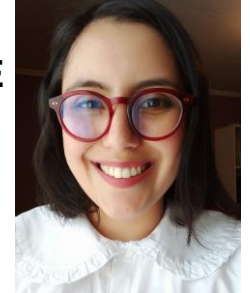


Tonino Costa



Marcelo Morán

LENGUAJE



Valentina Basso



Nicolás Pizarro

**CIENCIAS
SOCIALES**



Jacqueline Gaete



María Paz
Blanco



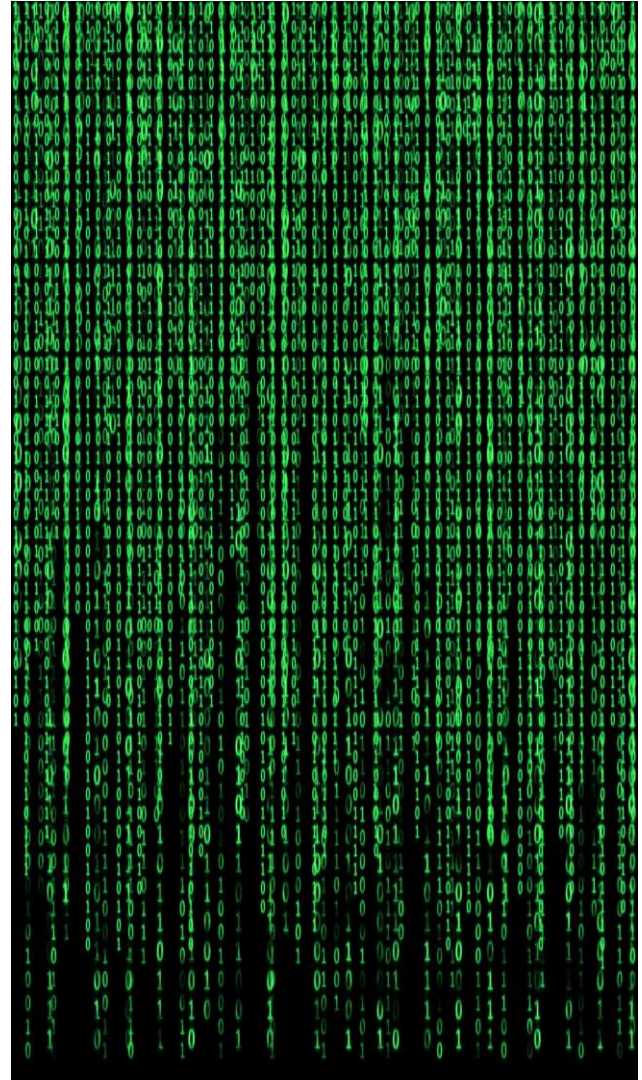
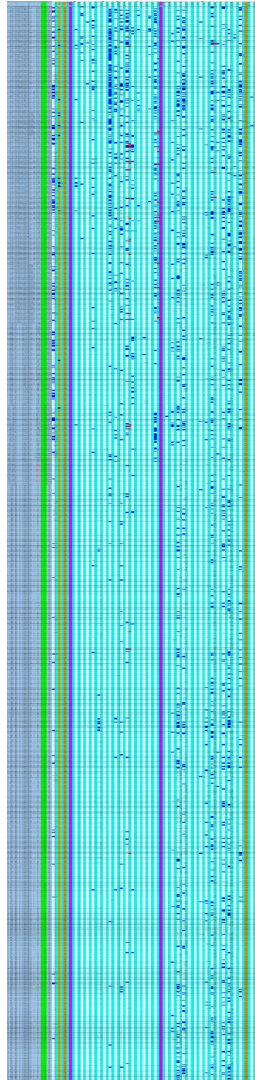
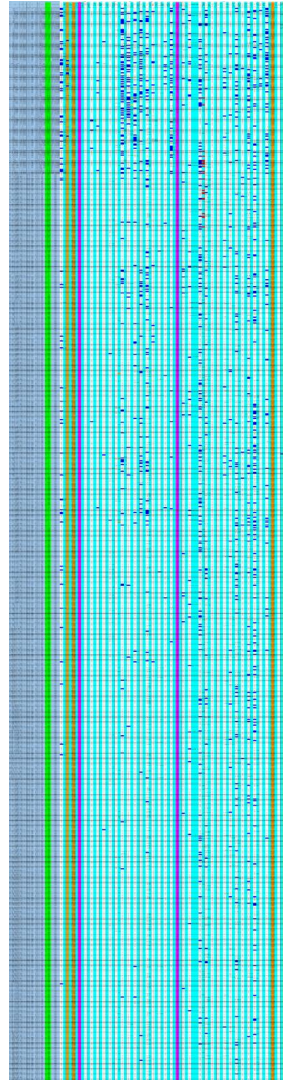
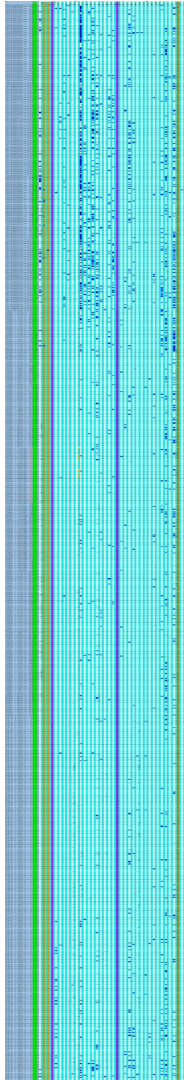
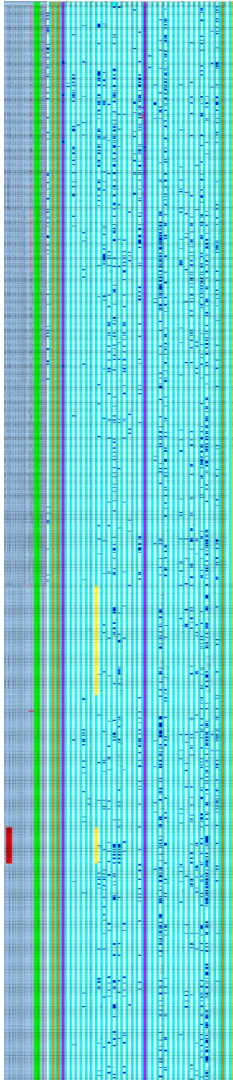
Paula Rada

CIENCIAS



Séverin Lions

Anotación de ítems: entre 1641 y 4140 hrs. de trabajo



Anotación de ítems: proyecciones de investigación

- Estudio de pistas hacia la opción correcta:
¿Are cues to the correct option reliable?

- Guía de construcción de ítems según disciplina
Toward discipline-dependent item-writing guidelines

- Diferentes disciplinas, mismas imperfecciones
How do the same item-writing flaws look like in different disciplines

Validación de SARI:

1 estudio de **utilidad**, 32 participantes, 340 ítems evaluados

1 estudio de **eficiencia**, 19 participantes, 40 ítems evaluados

		Disciplina				
		Ciencias	Historia	Lenguaje	Matemática	Todas
Número de participantes		9	6	8	9	32
Experiencia (años)		7,5	7,5	9,0	13,4	9,5
Facilidad de uso		4,3	4,5	4,5	4,0	4,3
Accesibilidad y Legibilidad (de 1 a 5)	Eficiencia	Ciencias		Lenguaje	Todos	
		Efecto facilitador		3,8	3,3	3,6
	Aporte exhaustividad		3,6	3,1	3,4	
	Aporte precisión		3,7	3,2	3,5	
	Ganancia tiempo		3,7	3,6	3,6	
	Eficiencia potencial		3,9	4,0	3,9	
	Total		3,7	3,4	3,6	
	Satisfacción asistencia		3,7	3,3	3,5	
	Utilidad asistencia		3,9	3,9	3,9	
	Eficacia (de 1 a 5)	Disposición para uso futuro		3,4	3,2	2,9
Recomendable		3,1	2,8	2,9	2,7	
Satisfacción general		2,7	2,3	2,4	2,4	
Total		3,1	2,8	2,7	2,6	
Satisfacción (de 1 a 5)		3,1	2,8	2,7	2,6	2,8
Todas las dimensiones		3,2	3,3	3,0	3,1	3,1

Validación de SARI:

1 estudio de **eficiencia**, 19 participantes, 40 ítems evaluados

	Claro		Exhaustivo		Preciso	
	Con SARI	Sin SARI	Con SARI	Sin SARI	Con SARI	Sin SARI
Ciencias	4,3	4,4	3,9	4,2	3,9	4,1
Lenguaje	3,9	3,6	3,8	3,9	4,0	3,9
Total	4,1	4,0	3,9	4,1	4,0	4,0

		Duración (min:seg)				
		Con SARI	Sin SARI (norm.)	Sin SARI	Delta (norm.)	Delta
Ciencias	Comisión 1	05:51	08:20	06:01	02:29	00:09
	Comisión 2	06:05	08:50	07:38	02:46	01:33
	Comisión 3	06:00	09:15	08:05	03:15	02:05
	Todas las comisiones	05:59	08:49	07:15	02:50	01:16
Lenguaje	Comisión 1	07:34	10:25	09:00	02:51	01:26
	Comisión 2	06:11	07:28	05:51	01:16	00:21
	Comisión 3	07:16	09:01	07:38	01:45	00:22
	Todas las comisiones	07:00	08:58	07:30	01:57	00:29
Todas asignaturas y comisiones		06:30	08:53	07:22	02:24	00:53

Validación de SARI: 3 ventajas de SARI

- Revisa sistemáticamente todos los criterios que son programados,
- Entrega una retroalimentación mucho más rápidamente,
- Es capaz de verificar los cambios realizados.

Por tanto, **permite asegurar la calidad de las preguntas, obtener esta calidad en un tiempo más reducido, y, en definitiva, ahorrar costos.**

Muchas perspectivas a **corto** y **largo** término

- Detección de imperfecciones de ensamblaje
- Procesamiento de imágenes y ecuaciones
- Predicción de datos psicométricos
- Detección de sesgos (cultural, territorial, de género...)
- **SARI 2.0: masivo y con impacto formativo**
- **Internacionalización**



Algoritmos



Interfaz



Estudios



Gestión


FONDEF SARI: ¡gracias!



**Subsecretaría de
Educación Superior**
Ministerio de Educación



**Agencia de
Calidad de la
Educación**



CiAE
CENTRO DE INVESTIGACIÓN
AVANZADA EN EDUCACIÓN



UNIVERSIDAD DE CHILE
INSTITUTO DE ESTUDIOS
AVANZADOS EN EDUCACIÓN **IE**



PREGUNTAS

**Si quieres trabajar con nosotros en estos
proyectos:**

Séverin Lions (severin.lions@ciae.uchile.cl)