



Universidad del País Vasco
Euskal Herriko Unibertsitatea

HiTZ

Hizkuntza Teknologiko Zentroa
Basque Center for Language Technology

www.hitz.eus

@hitz_zentroa



La tecnología del lenguaje, la próxima frontera de la Inteligencia Artificial

Informe anual
2020

INTRODUCCIÓN



2

La comprensión del lenguaje es uno de los principales objetivos de la inteligencia artificial y, aunque se ha avanzado mucho en el aprendizaje profundo, sigue siendo una de las principales limitaciones de la inteligencia artificial, ya que los ordenadores no serán inteligentes hasta que no entiendan el lenguaje. Para hacer frente a los nuevos desafíos y oportunidades de la transformación digital y la inteligencia artificial, es primordial fomentar la investigación en las tecnologías del lenguaje. Para ello, hemos creado HiTZ, con un doble propósito: ser un referente internacional en la investigación de las tecnologías del lenguaje, y el referente en el



tratamiento computacional del euskera. La actividad del centro comprende cinco líneas de actuación: formación, investigación, innovación y transferencia, fomentando nuestra cultura y el impacto en la sociedad. Este extenso ámbito requiere un amplio equipo interdisciplinar, que reúne a informáticos, lingüistas, ingenieros, traductores y sociólogos, entre otros. El trabajo de los investigadores de Aholab e Ixa, los dos grupos de investigación que conforman el centro HiTZ, dio sus frutos este año. Son de subrayar la creación con éxito de los primeros modelos de lenguaje neuronales para el euskera, el liderazgo del proyecto ELE que va a definir la es-

trategia sobre tecnología del lenguaje en Europa, los avances en la infraestructura europea CLARIN, y los 7 premios recibidos en los mejores congresos internacionales y en varias competiciones científicas, incluida la que impulsó el gobierno de Estados Unidos sobre la COVID. Esta sólida base nos permite continuar con energías renovadas y, en colaboración con centros tecnológicos, empresas e instituciones, convertir al País Vasco en un centro internacional de tecnologías del lenguaje e inteligencia artificial.

Eneko Agirre (Director del HiTZ) y German Rigau (Subdirector del HiTZ)

HiTZ EN CIFRAS

40

proyectos

22

contratos de
transferencia

71

publicaciones

3

tesis

4

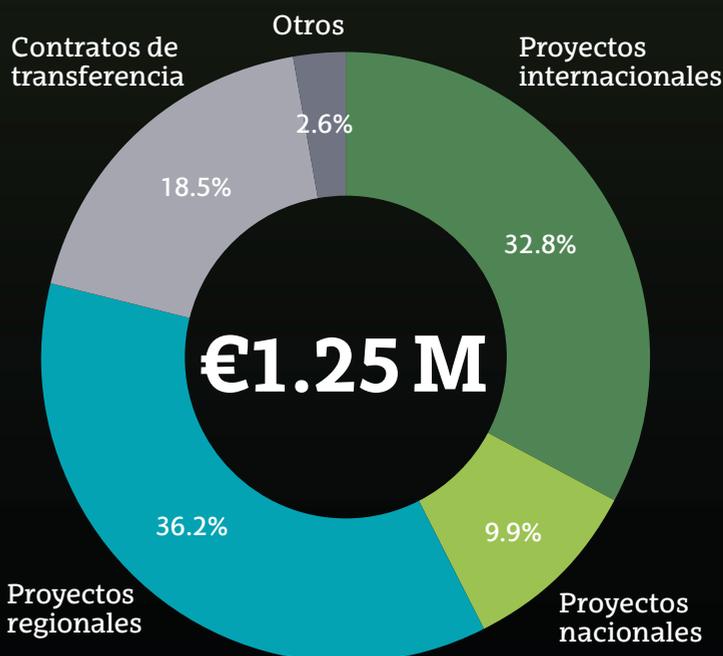


Miembros — 106

	Gestión y administración	3
	Profesores	39
	Becas de doctorado	13
	Investigadores postdoctorales	4
	Ramón y Cajal	1
	Estudiantes de HAP/LAP	10
	Estudiantes de EMLTC	2
	Estudiantes de doctorado	31
	Tesis doctorales defendidas	3



Presupuesto —



ORGANIZACIÓN

HiTZ es un centro de investigación multidisciplinar especializado en la **Inteligencia Artificial centrada en el lenguaje**, formado por miembros de 7 departamentos de la Universidad del País Vasco/Euskal Herriko Unibertsitatea. El objetivo del centro es **investigar** en el lenguaje y las tecnologías del habla, con un significativo esfuerzo en la **transferencia** de conocimiento y tecnología a las empresas. HiTZ está compuesto por dos grupos de investigación, Aholab e Ixa. Ambos poseen una amplia experiencia en investigación de base, creando recursos y herramientas lingüísticas y habiendo lanzado varios productos al mercado. HiTZ es **miembro fundador** de **CLARIN-K Center**, miembro de **CLAIRE** y miembro de pleno derecho de **BDVA** y **DAIRO**. A través de CLAIRE y BDVA, participamos en el European Partnership on Artificial Intelligence, Data and **Robotics**.



Eneko Agirre

Director



German Rigau

Subdirector



Ander Salaberria

Miembro de la Junta



Maite Oronoz

Miembro de la Junta

Los miembros del centro son referentes internacionales en sus áreas científicas. En este momento HiTZ está formado por **más de 60 miembros** que incluyen informáticos, lingüistas y 3 técnicos de investigación. En el último lustro, los investigadores que forman el centro han publicado más de 200 **publicaciones**. El grupo es líder aplicando técnicas de aprendizaje profundo y, en los dos últimos años los trabajos realizados en éste área han sido **citados en más de 4000 ocasiones**. Los miembros del centro han realizado labores de asesoramiento en Plan Nacional de Tecnologías del Lenguaje y actualmente están asesorando al Gobierno Vasco en labores similares.

**CLARIN
K CENTRE**



CLAIRE



BDVA BIG DATA VALUE ASSOCIATION



ahō LAB



**Kike
Fernandez**

Miembro de
la Junta



**Itziar
Aldabe**

Miembro de
la Junta



**Inma
Hernaez**

Miembro de
la Junta



**Aitor
Soroa**

Miembro de
la Junta

Tanto Ixa como Aholab han sido denominados grupos de alto rendimiento en la última evaluación realizada por el Consejo Científico del Gobierno Vasco. A lo largo de su historia, los grupos han participado en más de 200 **proyectos de investigación**, desde regionales hasta europeos. Además de participar en más de 100 **contratos industriales** con el objetivo de transferir tecnología a la industria.

HiTZ es miembro de **Erasmus Mundus+ European Masters Program in Language and Communication Technologies (LCT)**, un programa diseñado para aunar las demandas tanto de la industria como de la investigación en un campo de rápido crecimiento como es el de

la tecnología del lenguaje. HiTZ ofrece además un **programa Doctoral** en Procesamiento del Lenguaje Natural.

La Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU) es la principal entidad para la enseñanza e investigación en el País Vasco. La UPV/EHU está entre las 400 mejores universidades del mundo según el ranking Shangai, y el Gobierno de España la ha reconocido como Campus de Excelencia. La Universidad del País Vasco es una institución de más de 30 años de antigüedad, con 45.000 alumnos, 5.000 académicos de nivel mundial e instalaciones de vanguardia, distribuidas en 20 centros en sus 3 campus.

ÁREAS DE INVESTIGACIÓN



Investigador/a principal:

Extracción y recuperación de información



Aitor Soroa

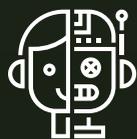


Investigador/a principal:

Traducción automática



Gorka Labaka



Investigador/a principal:

Interacción persona-computadora



Eneko Agirre



Investigador/a principal:

Recursos del habla y el lenguaje



Ainara Estarrona

INFRAESTRUCTURA

14

servidores multiprocesadores GNU/Linux

4

servidores de SPARC Solaris

1

HPC Cluster con 128 núcleos

 Si haces click en este símbolo podrás obtener más información



Análisis de texto



Investigador/a principal:



Rodrigo Agerri



Tecnologías de Voz



Investigador/a principal:



Inma Hernaez



Dominios médico y legal



Investigador/a principal:



Arantza Casillas



Humanidades digitales y educación



Investigador/a principal:



Mikel Iruskietia

1

sala aislada acústicamente con equipo de audio para grabaciones profesionales

7 servidores con

31

GPUs

Más de

260 TB

de capacidad de almacenamiento bruto en red

INVESTIGACIÓN Y TRANSFERENCIA



40

proyectos de
investigación

22

proyectos de
transferencia de
conocimiento

7

otros
proyectos

3

tesis
doctorales
defendidas

29

artículos de
revistas
(8 Q1)

38

artículos en
congresos
(8 A o A+)

1

capítulo
de libro

3

libros

13

becas de
doctorado





12 
estudiantes en
el Máster

31 
estudiantes en
el programa de
doctorado

PRÁCTICAS

2 
TFMs de EMLCT
finalizadas

10 
TFMs de HAP/
LAP finalizadas

91 
estudiantes en 2
cursos complemen-
tarios de Aprendiza-
je Profundo

3
Ikasiker

8
prácticas internas
y externas

ACTIVIDADES



25

seminarios



3

talleres

7

premios

Google Faculty Research Awards

eHealth-2020 Tournament

CodiEsp-D. Main evaluation metric:
Mean Average Precision

Ganador del Capitel@IberLEF 2020
shared task on Named Entity
Recognition (NER) for Spanish

COVID-19 Open Research Dataset
Challenge (CORD-19)

Mención de honor por mejor artículo en
EMNLP (Conference on Empirical
Methods in Natural Language Processing).

Premio "Outstanding Paper" en
COLING (International Conference
on Computational Linguistics).