
Artículo

[Alberto Fuentes](#) · 30 ene, 2023 · Lectura de 3 min

[Open Exchange](#)

Demo de OCR

OCR DEMO

Esta es una demo de la funcionalidad OCR utilizando la librería pero-ocr de Python.

Utilizaremos la librería en una instancia InterSystems IRIS.

Demo

Este es un ejemplo de los datos de entrada:

IN CONGRESS, JULY 4, 1776.

The unanimous Declaration of the thirteen united States of America.

When in the Course of human events, it becomes necessary for one people to dissolve the political bands which have connected them with another, and to assume among the powers of the earth, the separate and equal station to which the Laws of Nature and of Nature's God entitle them, a decent respect to the opinions of mankind requires that they should declare the causes which impel them to the separation. — We hold these truths to be self-evident, that all men are created equal, that they are endowed by their Creator with certain unalienable Rights, that among these are Life, Liberty and the pursuit of Happiness. — That to secure these rights, Governments are instituted among Men, deriving their just powers from the consent of the governed, — That whenever any Form of Government becomes destructive of these ends, it is the Right of the People to alter or to abolish it, and to institute new Government, laying its foundation on such Principles and organizing its Powers in such form, as to them shall seem most likely to effect their Safety and Happiness. — Prudence, indeed, will dictate that Governments long established should not be changed for light and transient causes; and accordingly all experience hath shewn, that mankind are more disposed to suffer, while evils are sufferable, than to right themselves by abolishing the forms to which they are accustomed. — But when a long train of abuses and usurpations, pursuing invariably the same Object evinces a design to reduce them under absolute Despotism, it is their right, it is their duty, to throw off such Government, and to provide new Guards for their future security. — Such has been the patient Sufferance of these Colonies; and such is now the necessity which constrains them to alter their former Systems of Government. — The history of the present King of Great Britain is a history of repeated injuries and usurpations, all having in direct and plain the establishment of an absolute Tyranny over these States. — To prove this, let Facts be submitted to a candid world. — He has refused his Assent to Laws, the most wholesome and necessary for the public good. — He has forbidden his Governors to pass Laws of immediate and pressing importance, unless suspended in their operation till his Assent should be obtained; and when so suspended, he has utterly neglected to attend to them. — He has refused to pass other Laws for the accommodation of large districts of People, unless those people would relinquish the right of Representation in the Legislature, a right inestimable to them and formidable to tyrants only. — He has called together legislative Bodies at places unusual, uncomfortable, and distant from the depository of their Public Records, for the sole purpose of fatiguing them into compliance with his measures. — He has dissolved Representative Houses repeatedly, for opposing with manly firmness his invasions on the rights of the People. — He has refused for a long time, after such dissolutions, to cause others to be elected; whereby the Legislative Powers, incapable of Annihilation, have returned to the People at large for their exercise; the State remaining in the mean time exposed to all the dangers of Invasion from without, and convulsions within. — He has endeavoured to prevent the Population of these States; for that purpose obstructing the Law for Naturalization of Foreigners; refusing to pass others to encourage their migrations hither, and raising the conditions of new Appropriations of Lands. — He has obstructed the Administration of Justice, by refusing his Assent to Laws for establishing Judiciary Powers. — He has made Judges dependent on his Will alone, for the tenure of their Offices, and the amount and payment of their salaries. — He has created a multitude of New Offices, and sent hither swarms of Officers to harass our People, and eat out their substance. — He has kept among us, in Times of Peace, standing Armies without the Consent of our Legislature. — He has affected to render the Military independent of and superior to the Civil Power. — He has combined with us to subject us to a Jurisdiction foreign to our Constitution, and unacknowledged by our Laws; giving his Assent to their Acts of pretended Legislation: — For Quarters large bodies of armed troops among us: — For protecting them, by a mock Trial, from punishment for any Murders which they should commit on the Inhabitants of these States: — For cutting off our Trade with all parts of the world: — For imposing Taxes on us without our Consent: — For depriving us in many cases, of the benefits of Trial by Jury: — For transporting us beyond Seas to be tried for pretended offences: — For abolishing the free System of English Laws in a neighbouring Province, establishing therein an Arbitrary government, and enlarging its Boundaries so as to render it at once an example and fit instrument for introducing the same absolute rule into these Colonies: — For taking away our Charters, abolishing our most valuable Laws, and altering fundamentally the Forms of our Governments: — For suspending our own Legislatures, and declaring themselves invested with powers to legislate for us in all cases whatsoever. — He has abdicated Government here, by declaring us out of his Protection and waging War against us. — He has plundered our seas, ravaged our Coasts, burnt our towns, and destroyed the lives of our People. — He is at this time transporting large Armies of foreign Mercenaries to compleat the works of death, desolation and tyranny, already begun with circumstances of cruelty & perfidy scarcely paralleled in the most barbarous ages, and totally unworthy the Head of a civilized nation. — He has constrained our fellow Citizens taken Captive on the high seas to bear Arms against their Country, to become the executioners of their friends and Brethren, or to fall themselves by their Hands. — He has excited domestic insurrections amongst us, and has endeavoured to bring on the Inhabitants of our frontiers, the merciless Indian Savages, whose known rule of warfare, is an undistinguished destruction of all ages, sexes and conditions. — In every stage of these Oppressions We have Petitioned for Redress in the most humble terms: — Our repeated Petitions have been answered by insult and injury. — A Prince, whose character is thus marked by every act which may define a Tyrant, is unfit to be the ruler of a free People. — Nor have We been wanting in attentions to our British brethren. We have warned them from time to time of attempts by their Legislature to extend an unwarrantable Jurisdiction over us. We have reminded them of the circumstances of our emigration and settlement here. We have appealed to their native justice and magnanimity, and we have conjured them by the ties of our common kindred to disavow these usurpations, which would inevitably interrupt our connections and correspondence. — They too have been deaf to the voice of justice and of concinnity. — We must, therefore, acquiesce in the necessity, which denounces our Separation, and hold them, as we hold the rest of mankind, Enemies in War, in Peace Friends. — We, therefore, the Representatives of the united States of America, in General Congress, Assembled, appealing to the Supreme Judge of the world for the rectitude of our intentions, do, in the Name, and by Authority of the good People of these Colonies, solemnly publish and declare, That these United Colonies are, and of Right ought to be, Free and Independent States; that they are Absolved from all Allegiance to the British Crown, and that all political connection between them and the State of Great Britain, is and ought to be totally dissolved; and that as Free and Independent States, they have full Power to levy War, conclude Peace, contract Alliances, establish Commerce, and to do all other Acts and Things which Independent States may of right do. — And for the support of this Declaration, with a firm reliance on the Protection of Divine Providence, we mutually pledge to each other our Lives, our Fortunes and our sacred Honor.

Butler, Gwinnett
Lynch, Hall
Geo. Walton.

John Hancock
Joseph B. Beeves
John Penn

Edward Rutledge

Thos. Mifflin
Thos. Mifflin
Arthur Middleton

Samuel Adams
Wm. Paine
Thos. Paine
Charles Carroll of Carrollton

George Wythe
Richard Henry Lee
Thos. Jefferson
Wm. Harrison
Thos. Nelson
Charles Braxton

John Morris
Benjamin Rush
Benjamin Franklin
John Morton
Geo. Taylor
James Wilson
Gt. Twp
Cesar Rodney
Gt. Twp
Thos. Mifflin

Rich. Stockton
Wm. Livingston
John Dickinson
John Hart
Abra. Clark

John Jay
John Jay
John Jay
John Jay

John Jay
John Jay
John Jay
John Jay

John Bartlett
Wm. Mifflin
John Adams

John Adams
John Adams

John Adams
John Adams

John Adams
John Adams
John Adams
John Adams

Y este es el resultado del OCR, donde tenemos la siguiente información:

- El texto, que está en la etiqueta TextEquiv
- La confianza de la lectura, presente en el atributo conf de la etiqueta TextEquiv
- Las coordenadas del texto, que están en la etiqueta Coords

<PcGts xmlns="http://schema.primaresearch.org/PAGE/gts/pagecontent/2019-07-15" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://schema.pri

```

maresearch.org/PAGE/gts/pagecontent/2019-07-15/pagecontent.xsd">
  <Metadata>
    <Creator>Pero OCR</Creator>
    <Created>2022-12-13T08:47:12.207893+00:00</Created>
    <LastChange>2022-12-13T08:47:12.207893+00:00</LastChange>
  </Metadata>
  <Page imageFilename="/irisdev/app/misc/in/United_States_Declaration_of_Independence.jpg" imageWidth="3923" imageHeight="4656">
    <TextRegion id="r004">
      <Coords points="977,121 932,121 932,283 1189,277 1178,114 977,121"/>
      <TextLine id="r004-l002" index="0" custom="heights_v2:[123.7,39.0]">
        <Coords points="932,121 932,283 1189,277 1178,114 1130,118 1100,117 1048,121 1019,121 977,121 932,121"/>
        <Baseline points="932,244 977,244 1019,244 1058,244 1100,241 1138,241 1186,238"/>
        <TextEquiv conf="1.000">
          <Unicode>IN</Unicode>
        </TextEquiv>
      </TextLine>
    </TextRegion>
    <TextRegion id="r008">
      <Coords points="1461,107 1241,107 1241,287 2116,290 2549,281 2986,287 3211,277 3203,98 2978,107 2552,101 2114,111 1461,107"/>
      <TextLine id="r008-l001" index="0" custom="heights_v2:[133.8,45.8]">
        <Coords points="1241,107 1241,287 1464,287 1681,284 1896,287 2116,290 2549,281 2986,287 3211,277 3203,98 2978,107 2552,101 2327,107 2114,111 1461,107 1241,107"/>
        <Baseline points="1241,241 1463,241 1682,238 1897,241 2116,244 2331,241 2550,235 2765,238 2984,241 3209,231"/>
        <TextEquiv conf="0.910">
          <Unicode>CONGRESS, JULY 4, 1776.</Unicode>
        </TextEquiv>
      </TextLine>
    </TextRegion>
    <TextRegion id="r001">
      <Coords points="591,452 208,455 209,611 1723,602 2100,608 2855,595 3617,598 3618,442 2855,439 2098,452 1725,446 591,452"/>
      <TextLine id="r001-l003" index="0" custom="heights_v2:[113.8,42.1]">
        <Coords points="208,455 209,611 592,608 968,605 1218,605 1723,602 2100,608 2855,595 3231,595 3617,598 3618,442 2855,439 2477,446 2098,452 1725,446 1346,449 1159,449 591,452 208,455"/>
        <Baseline points="209,569 592,566 968,563 1347,563 1723,559 2100,566 2479,559 2855,553 3231,553 3617,556"/>
        <TextEquiv conf="0.289">
          <Unicode>Dhe unaniwons Declaratton of te Heten maiss States of TNmerica</Unicode>
        </TextEquiv>
      </TextLine>
    </TextRegion>
    <TextRegion id="r000">
      <Coords points="161,3474 161,3539 423,3539 633,3536 958,3468 1354,3472 1764,3581 1944,3571 2246,3730 2653,3718 2718,3633 2855,3608 3238,3557 3514,3554 3739,3462 3762,3397 3761,3333 3735,3139 3743,2696 3730,2177 3735,1920 3723,1727 3736,1341 3717,1014 3724,943 3719,818 3697,693 1012,712 795,776 421,776 183,786 173,848 168,1032 173,2392 160,2713 161,3474"/>
      <TextLine id="r000-l004" index="0" custom="heights_v2:[40.2,18.4]">
        <Coords points="2796,699 2501,703 2205,706 1909,709 1611,709 1460,709 1012,712 1013,771 1315,768 1611,768 1910,768 2206,764 2502,761 2800,758 3096,758 3392,755 3698,752 3697,693 3392,696 3096,699 2802,699 2796,699"/>
        <Baseline points="1013,752 1315,749 1611,749 1910,749 2206,746 2501,743 2800,

```

```

739 3096,739 3392,736 3697,733"/>
    <TextEquiv conf="0.281">
        <Unicode>hen n lí loune z human venl, i kemu nematy k mpeop? toíohohhehtca
l bandí uhich have connechdí tem vith ancthet, andíl</Unicode>
    </TextEquiv>
</TextLine>
<TextLine id="r000-l005" index="1" custom="heights_v2:[40.4,19.2]">
    <Coords points="421,776 421,836 794,836 1157,839 1521,839 1884,836 2244,833 2
608,829 2971,826 3335,823 3704,813 3703,754 3333,763 2970,767 2607,770 2244,773 1884,
776 1520,779 1157,779 795,776 421,776"/>
    <Baseline points="421,817 794,817 1157,820 1521,820 1884,817 2244,813 2608,81
0 2971,807 3334,804 3704,794"/>
    <TextEquiv conf="0.211">
        <Unicode>o hi ?hwes f he eail, fie rehatal andequal flohon & ufch lhe l
aav . kalut and Aloil ped entilt ttem, a dant rafech to the ofunin o manknd tequie
s fhat thep</Unicode>
    </TextEquiv>
</TextLine>
<!-- Truncated for readme-->
<TextLine id="r019-l105" index="9" custom="heights_v2:[47.5,21.7]">
    <Coords points="3209,4269 3214,4336 3257,4333 3296,4333 3332,4333 3373,4337 3
411,4337 3452,4336 3488,4333 3527,4333 3575,4333 3575,4264 3527,4264 3488,4264 3446,4
267 3411,4267 3373,4267 3338,4264 3296,4264 3257,4264 3209,4267 3209,4269"/>
    <Baseline points="3212,4315 3257,4312 3296,4312 3334,4312 3373,4315 3411,4315
3450,4315 3488,4312 3527,4312 3575,4312"/>
    <TextEquiv conf="0.283">
        <Unicode>imuiaa</Unicode>
    </TextEquiv>
</TextLine>
<TextLine id="r019-l108" index="10" custom="heights_v2:[40.9,20.2]">
    <Coords points="3198,4406 3251,4409 3294,4409 3339,4412 3394,4415 3398,4354 3
343,4351 3298,4348 3251,4348 3202,4345 3198,4406"/>
    <Baseline points="3199,4386 3251,4389 3296,4389 3341,4392 3395,4395"/>
    <TextEquiv conf="0.370">
        <Unicode>Qlver</Unicode>
    </TextEquiv>
</TextLine>
<TextLine id="r019-l114" index="11" custom="heights_v2:[39.2,19.1]">
    <Coords points="3181,4469 3225,4466 3263,4466 3305,4466 3327,4466 3352,4465 3
386,4462 3424,4462 3472,4462 3472,4404 3424,4404 3386,4404 3341,4407 3305,4407 3263,4
407 3225,4407 3177,4411 3181,4469"/>
    <Baseline points="3180,4450 3225,4447 3263,4447 3305,4447 3344,4447 3386,4443
3424,4443 3472,4443"/>
    <TextEquiv conf="0.312">
        <Unicode>Vbal?ew</Unicode>
    </TextEquiv>
</TextLine>
</TextRegion>
<TextRegion id="r020">
    <Coords points="3685,4399 3607,4399 3607,4465 3685,4465 3685,4399"/>
    <TextLine id="r020-l112" index="0" custom="heights_v2:[44.3,21.4]">
        <Coords points="3607,4399 3607,4465 3685,4465 3685,4399 3607,4399"/>
        <Baseline points="3607,4443 3685,4443"/>
        <TextEquiv conf="0.486">
            <Unicode>17.</Unicode>
        </TextEquiv>
    </TextLine>
</TextRegion>
</Page>

```

</PcGts>

Instalación

```
git clone <this repo>
```

¡Importante! Esta demo requiere instalar dos modelos (ambos son requeridos)

Para instalar un modelo, descárgalo desde la página de descargas y descomprímelo en el directorio misc/pero-ocr-fix-computation-on-cpu del proyecto:

- <https://github.com/grongierisc/iris-pero-ocr/releases/download/v1.0.0/OC...>
- <https://github.com/grongierisc/iris-pero-ocr/releases/download/v1.0.0/Pa...>

Después, ejecuta:

```
docker-compose up
```

Uso

Coloca cualquier imagen de muestra de la carpeta samples y cópiala en la carpeta misc/in para que sea procesada por el OCR.

Los resultados estarán en la carpeta misc/out.

Verás los archivos xml con los resultados y las imágenes con el texto detectado.

Puedes monitorizar el progreso [aquí](#) y accediendo con `_SYSTEM` y `SYS`.

Cómo funciona en IRIS

El OCR es un Business Service que analiza todos los ficheros en la carpeta misc/in y pone los resultados en una cola de mensajes.

La cola de mensajes es consumida por una Business Operation que pone los resultados en la carpeta misc/out.

El código está en la carpeta src/python/pero-ocr.

[#Python #InterSystems IRIS](#)

[Ir a la aplicación en InterSystems Open Exchange](#)

URL de fuente: <https://es.community.intersystems.com/post/demo-de-ocr>