

# EndecaCommunity.com

Everything you want to know about Endeca

Search

or

Ask a Question

## Endeca Community Solutions

CAS



Dear Endeca Community user,

On July 31, 2015 we closed EndecaCommunity.com, and asked users to instead take advantage of the collaboration tools and communities now provided by Oracle.

Why we did this? Because EndecaCommunity.com has fulfilled its purpose.

We created EndecaCommunity.com to fill a temporary void, to support teams working with Endeca technology in both Commerce and Data Insight capacities. The intention was to provide continuity through the transition period that followed Oracle's acquisition of Endeca, until Oracle had ramped up its own Endeca support community.

Now that Oracle's Endeca communities and formal user groups are well established, it's better for you as an Endeca user to find all the support you need from OAUG Special Interest Group and Oracle sources, rather than split your time having to cross-post to multiple disparate sites to find the answers you need.

We recommend joining the [Endeca Community on LinkedIn](#) or the [OAUG Endeca Special Interest Group](#). More support for Endeca can be found on the Oracle community website at [Technical Questions](#) and [Endeca Information Discovery](#).

On the following pages you will find the most frequently asked questions about CAS on EndecaCommunity.com and the best solutions as determined by members of the Community.

Thank you for taking advantage of EndecaCommunity.com and for trusting RealDecoy as your Endeca authority.

Sincerely,

Richard Isaac  
President & CEO, RealDecoy

# Table of Contents

## CAS

- **Problem with recordstore-cmd utility in partial update** 4
- **Exclude text from CAS webcrawl** 5

# CAS

## Problem with recordstore-cmd utility in partial update

**Q:** I have this line:

```
"call %CAS_ROOT%\bin\recordstore-cmd.bat read-baseline -a OfferPruebas -f  
C:\Endeca\Apps\Discover\test_data\partial\partial_data.xml"
```

in my partial\_update.bat in order to transform my csv input (horizontal comma separated) to xml indexable by the partial update process. But when my csv has a record with the Endeca.Action = "DELETE" the recordstore-cmd utility ignores it and doesn't transform it to xml record.

How is this possible? Any ideas?

**A:** I'm confused.

The content of the csv is this one for example:

```
record.id,record.type,offer.sell,offer.skuReposi,offer.list,offer.sale,offer.idShi,offer.amount,  
offer.idShipp,offer.amoB,offer.idingC,offer.apC,offer.sv,offer.rg,offer.prm,offer.resched,End  
eca.Action 12,offer de Manu lanzando prueba fin de  
semana,43578,12345678901,999,888,124,7,125,6,126,3,40,1,4,1,UPSERT 13,,,,,,,,,,,,,1,,14,DELETE  
14,offer,41090,3452543441877,999,888,166,8,121,462,10564,1,10,1,4,1,UPSERT
```

Here you mentioned that you are converting csv to xml but when I read following, it looks like you are actually converting CAS record-store to XML.

```
Call %CAS_ROOT%\bin\recordstore-cmd.bat read-baseline -a Ofertas2 -f  
C:\Endeca\Apps\Discover\test_data\partial\partial_offers_data.xml
```

As you have UPSERT and DELETE "stuff", I am assuming you are doing the following: using CAS crawl to create CAS record store from your input CSV, and then using above mentioned CAS command to convert it to XML which is then fetched to the pipeline.

If so, you need not do this. Once you have created the CAS record store, you can directly fetch it to endeca (partial) pipeline and do the incremental index. Check the following blog for help with this:

<http://themayankbatratechy.blogspot.com/2014/07/endeca-partial-update-with-cas.html>

## Exclude text from CAS webcrawl

**Q:** Is it possible to have an option within CAS to exclude or convert parts of text on a website? For example I want to crawl a question on this website. The webcrawler will then include all text on the page like the ones on the right side in the yellow blocks like "Once you sign in you will be able to subscribe for any updates here" and adds them in Endeca Document\_Text.

But I only want the text from the question and the related answers. So is it possible to exclude/filter/convert the rest of the text out?

**A:** The short answer is "Yes, you can!"

If you look to this website's source HTML, you'll notice that the question is wrapped in a div class="question-body".

That's the portion on the webpage you'll want to extract, so if you look into your XML configuration file you'll find a property named 'document.prune.xpath', where you'll have to insert an xpath expression, like: `//*[@id='item-right']/div[1]`

You're saying that you want the first div after an element with the id 'item-right', which is the parent of the div that you want. Now all the pages crawled will look for that particular div and extract it.