

MASARYKOVA UNIVERZITA  
FAKULTA INFORMATIKY



# Protokol Memento

PV070 DIGITÁLNÍ KNIHOVNY

**Václav Blahut**

**UČO 395963**

Brno, listopad 2016

## Základní údaje

Název projektu: Protokol Memento

Nositelé: Research Library of the Los Alamos National Laboratory  
Computer Science Department of Old Dominion University

URL: <http://mementoweb.org/>

# 1. Úvod

Tento článek pojednává o protokolu Memento[1], který si dává za cíl umožnit uživateli webu přístup k historii webových stránek stejně jednoduchým způsobem, jako při prohlížení stránek aktuálních. V praxi jde pouze o definici jednoduchých pravidel a rozšíření protokolu HTTP s cílem sjednotit existující přístupy webových stránek a archivů k této problematice.

Projekt Memento je výsledkem spolupráce týmů z Research Library of the Los Alamos National Laboratory a Computer Science Department of Old Dominion University. První zveřejněné informace o projektu pochází z roku 2009[2], přičemž návrh protokolu završený předložením RFC 7089[3] byl dokončen v roce 2013. Ve stejném roce byl vydán i doplněk pro prohlížeč Chrome[4], který umožňuje jednoduše využívat tento protokol koncovým uživatelům. Od té doby se podpora protokolu webovými stránkami postupně rozšiřuje, velkou událostí roku 2016 bylo přidání podpory protokolu Memento na oficiálních webech W3C[2].

## 2. Cíle projektu Memento

Přístup k historii webových stránek mají uživatelé i bez protokolu Memento. Existují tzv. webové archivy, které pravidelně prochází a ukládají obsah vybraných webových stránek. Po zadání URL adresy požadované stránky má pak v případě, že se daný web v konkrétním archivu nachází, uživatel možnost zobrazit historii této stránky v různých časových okamžicích. Nejznámějším webových archivem je Internet Archive[5], v česku existuje Český Webarchiv[6], který však protokol Memento nepodporuje. Ač nejde o webový archiv jako takový (nelze vybrat libovolné datum hlouběji v historii), je zde nutno zmínit i systém Google Webcache[7], který je schopen poskytnout jen jeden pohled na obsah webových stránek, starý zpravidla v řádu týdnů. Oproti klasickým archivům má však opravdu masivní pokrytí Internetu a často je tak jedinou cestou k pohledu do historie webů, které nejsou sledovány webovými archivy.

Jak již bylo uvedeno, cílem projektu je definovat jednotný protokol pro přístup k historii webových stránek. Existují dva hlavní způsoby, jak lze tento přístup získat. Prvním, preferovaným, ovšem zdaleka ne vždy možným způsobem je to, kdy webová stránka dokáže poskytnout svoji historii sama o sobě. Toto je zpravidla možné a implementačně jednoduché u webů, které podporují verzování obsahu. Jako příklad lze

uvést systémy založené na publikačních systémech (Wordpress, Joomla a další), stránky typu Wiki (Wikipedia) nebo server GitHub. Informace o tom, z jakého časového okamžiku si uživatel přeje obsah webu zobrazit, je pak podle protokolu Memento předána pomocí speciálních parametrů v HTTP dotazu.

Druhým způsobem, jak získat pohled do historie webu, jsou již zmíněné webové archivy, které pravidelně prochází vybrané stránky a archivují jejich aktuální obsah. Protokol Memento počítá s oběma způsoby pohledu do historie a předkládá jednotnou definici, jak by měly weby a webové archivy tyto pohledy do historie podporovat.

### 3. Typické scénáře použití a definice

Existuje více scénářů použití tohoto protokolu. Tím základním a nejjednodušším je ten, že uživatel si prohlíží libovolnou webovou stránku (označovanou **URI-R**, tedy aktuální verzi) a z nějakého důvodu by se rád podíval na tuto stránku tak, jak vypadala v nějaký daný okamžik v minulosti. Ve svém prohlížeči tedy vybere datum a čas, který ho zajímá, a tuto volbu potvrdí. Prohlížeč (resp. jeho část zodpovědná za funkci protokolu Memento) následně kontaktuje server, na kterém je webová stránka umístěna s dotazem, zda je tento server schopen poskytnout historii dané webové stránky (URI-R) v daném čase. Pokud ano, server vrátí odkaz na obsah stránky v tomto čase (nazývaný **URI-M**, tedy Memento – pohled na stránku v historii) a tyto jsou uživateli zobrazeny. Pokud server historii webu nabízí, nemá ovšem k dispozici pohled v daném okamžiku, je nalezena verze časově nejbližší a uživatel je na toto následně upozorněn s informací, z jakého okamžiku daná verze pochází.

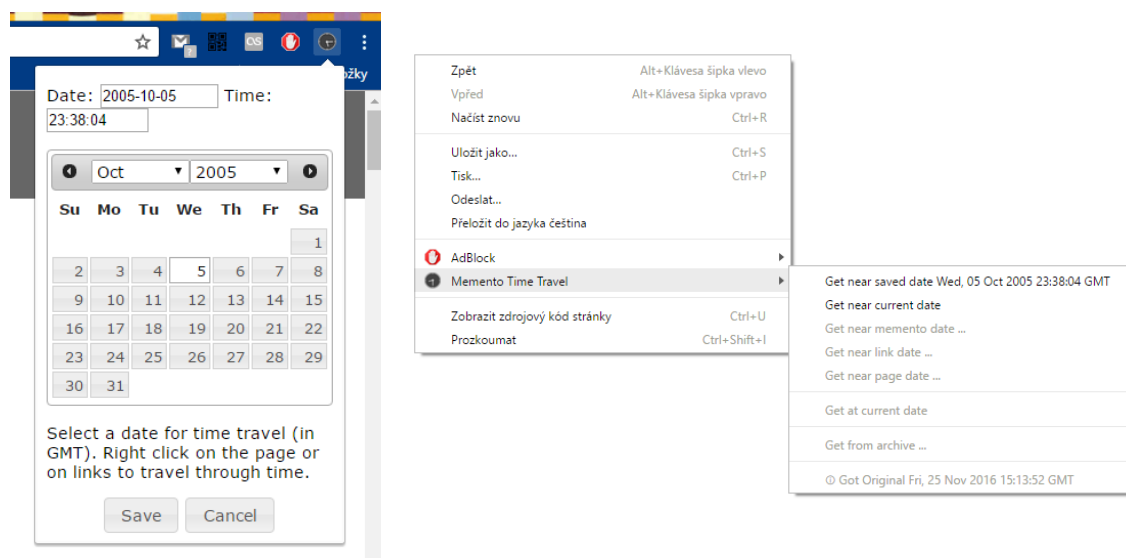
V případě, že server pohled do historie nenabízí vůbec, kontaktuje prohlížeč tzv. časovou bránu (označovanou **URI-G**, tedy TimeGate), což je speciální služba, která je zodpovědná za proces hledání ve webových archivech. V případě nalezení stránky v archivu je opět prohlížeči vrácen odkaz URI-M, pod nímž je k dispozici verze dané stránky v historii, a prohlížeč uživateli tuto stránku zobrazí s případným upozorněním, pokud čas neodpovídá uživatelem zadanému času.

Dalším scénářem může být to, že uživatel potřebuje všechny dostupné historické verze stránky, například aby si z nich vybral jednu konkrétní podle toho, co je k dispozici. V takovém případě se obdobně jako v prvním případě prohlížeč dotazuje pomocí HTTP nejdříve původního serveru ohledně URI-R s požadavkem na seznam všech dostupných

verzí, případně se stejně dotazuje i časové brány URI-G. Odpovědí mu pak je tzv. časová mapa (označovaná **URI-T**), tedy seznam odkazů URI-M spolu s odpovídajícími datумы. Formát časové mapy v protokolu Memento přímo definován není, jako příklad je však uveden široce využívaný JSON[8] nebo Link Format[9]. Pro případ rozsáhlých časových map, které obsahují příliš mnoho záznamů, se počítá i s indexováním a stránkováním časových map.

## 4. Praktické použití

Nejkomfortnější způsobem používání protokolu je doplněk pro prohlížeč Chrome. Po stažení a instalaci se uživateli na horním panelu zobrazí nové tlačítko, které po rozkliknutí umožňuje vybrat požadované datum a čas. Poté jen stačí kliknout pravým tlačítkem kdekoliv na zobrazené stránce a v podmenu Memento Time Travel vybrat první volbu. Doplněk poté popsaným způsobem kontaktuje servery a uživatele v případě úspěchu přesměruje přímo na historickou verzi stránky s případným varováním, pokud datum a čas přesně neseďí. Podobně fungují i dedikované prohlížeče pro systém Android.



Obrázek 1- Doplněk prohlížeče Chrome

Druhou, méně komfortní možností, která však nevyžaduje instalaci žádných doplňků nebo softwaru, je návštěva webu <http://timetravel.mementoweb.org/>, kde uživatel vloží odkaz na danou stránku a vybere datum a čas. Po vyhledání je přesměrován na historickou verzi stránky stejně jako při použití doplňku. V tomto případě jde tedy spíše o jakýsi agregátor webových archivů, které se obvykle používají stejným způsobem.

## 5. Rozšíření protokolu HTTP pro Memento

Jelikož protokol Memento pracuje čistě na úrovni protokolu HTTP (HTTPS), a to konkrétně jen s dotazy typu GET a HEAD, pro zahrnutí informací o požadovaném datu a celé komunikaci tak, jak bylo popsáno, bylo potřeba definovat dvě nové HTTP hlavičky, a dále nové hodnoty pro již existující hlavičky. Nová hlavička *Accept-Datetime* nabývá hodnoty časové známky při dotazu a deklaruje požadovaný časový moment. Hlavička *Memento-Datetime* je pak obsažena v odpovědi a obsahuje časovou známku odpovídající reálnému momentu, z kterého konkrétní URI-M pochází.

Pro již existující hlavičku *Vary* byla přidána nová hodnota *accept-datetime*, která signalizuje v rámci celé komunikace možnost vyjednávání o datu a čase. Existující hlavička *Link* je rozšířena o hodnoty identifikující typ odkazu, odpovídající jednotlivým označením URI-R (*original*), URI-G (*timegate*), URI-T (*timemap*) a URI-M (*memento*). Další, již existující hodnoty jsou dále různě používány v kontextu jednotlivých situací během komunikace.

## 6. Zhodnocení a přínos

Motivace autorů protokolu Memento definovat jednotný systém pro přidání časové dimenze do protokolu http, a výrazně tak zjednodušit přístup k historii webových stránek je zřejmá a jejich snaha chvályhodná. Vzhledem k akademickému pozadí a celkové jednoduchosti protokolu jak co se změnil jak protokolu HTTP, tak i implementace konkrétními weby týče byla na začátku jistě velká naděje, že se tento protokol uchytí a začne být masověji používán jako běžná součást webu a webových prohlížečů. Ovšem ani 7 let od začátku projektu a 3 roky od vydání doplňku pro Chrome a předložení RFC k žádnému masovému rozšíření nedošlo. Vypovídajícím číslem budiž počet stažení zmíněného doplňku v obchodě Chrome Web Store, které je v současnosti jen lehce přes 3000, přičemž byl tento doplněk ohodnocen jen šesti uživateli.

V době psaní tohoto článku měl jeho autor reálnou motivaci protokol Memento využít. Na serveru smsticket.cz potřeboval zpětně zjistit cenu vstupného na již proběhlou kulturní akci. Server však po ukončení předprodeje vstupenek již cenu neukazuje a šlo tedy o ukázkový případ, kdy by bylo Mementa možno využít. Server smsticket.cz však přímo protokol Memento nepodporuje a zároveň ani není sledován žádným z kompatibilních webových archivů, k historické verzi se tedy autor s jeho použitím

nedostal. Při použití služby Google Webcache však byla historická verze stránky s kulturní akcí dostupná a autor tak cenu vstupného nakonec úspěšně zjistil. Ač tedy jde o zajímavý, užitečný a jednoduše implementovatelný protokol, k možnosti reálného použití má kvůli malému pokrytí reálného webu bohužel stále daleko.

## 7. Reference

- [1] *Memento Project* [online]. [cit. 2016-11-23]. Dostupné z: <http://mementoweb.org/about/>
- [2] *Memento at W3C* [online]. 2016 [cit. 2016-11-23]. Dostupné z: <https://www.w3.org/blog/2016/08/memento-at-the-w3c/>
- [3] SOMPEL, Herbert van de, Robert SANDERSON a Michael NELSON. HTTP Framework for Time-Based Access to Resource States – Memento. *Memento project* [online]. 2013 [cit. 2016-11-23]. Dostupné z: <http://mementoweb.org/guide/rfc/>
- [4] *Memento Time Travel*. [online]. 2013 [cit. 2016-11-23]. Dostupné z: <https://chrome.google.com/webstore/detail/memento-time-travel/jgbfpjledahoajcppakbgilmojkaghgm?hl=en&gl=US>
- [5] *Internet Archive* [online]. [cit. 2016-11-23]. Dostupné z: <https://archive.org/>
- [6] *Český Webarchiv* [online]. [cit. 2016-11-23]. Dostupné z: <http://webarchiv.cz/cs/>
- [7] *Google Webcache* [online]. [cit. 2016-11-23]. Dostupné z: <http://google.cz/>
- [8] *JSON* [online]. [cit. 2016-11-23]. Dostupné z: <http://json.org/>
- [9] Link Relations. *IANA Protocol Registries* [online]. [cit. 2016-11-23]. Dostupné z: <http://www.iana.org/assignments/link-relations/link-relations.xhtml>

## 8. Metadata

<dc:title>Protokol Memento</dc:title>  
<dc:creator>Václav Blahut</dc:creator>  
<dc:subject>Protokol Memento, Time Travel</dc:subject>  
<dc:description>Článek se zabývá protokolem Memento, který usnadňuje přístup k historickým verzím webových stránek.</dc:description>  
<dc:date>2016-11-23</dc:date>  
<dc:type>text</dc:type>  
<dc:language>cz</dc:language>  
<dc:format>application/pdf</dc:format>